Saliency Processing in the Human Brain

Dissertation

zur Erlangung des akademischen Grades Doktor rerum naturalium (Dr. rer. nat.)

im Fach Psychologie

eingereicht an der Mathematisch-Naturwissenschaftlichen Fakultät II

der Humboldt-Universität zu Berlin

von Dipl.-Psych. Carsten Bogler

Prof. Dr. Jan-Hendrik Olbertz	Prof. Dr. Elmar Kulke
Präsident	Dekan

Gutachter/Gutachterin:

Prof. Dr. John-Dylan Haynes

Prof. Dr. Philipp Sterzer

Prof. Dr. Peter König

Eingereicht am: 05.09.2013

Tag der Verteidigung: 23.06.2014

Abstract

The human visual system is exposed to a vast amount of information every second; a filter is needed to highlight relevant elements. Attention to visual stimuli can be guided by top-down search strategies, for example if we look for something green, or by bottom-up information, i.e. when something stands out from its background. The property of a specific position to stand out in a visual scene is referred to as saliency. On the neural level, a representation of a saliency map, i.e. a map that encodes the saliency for each position of the visual field, is assumed to exist. However, to date it is still unclear where such a representation is located in the brain.

This dissertation describes three experiments that investigated different aspects of bottom-up saliency processing in the human brain using functional magnetic resonance imaging (fMRI). Neural responses to different salient stimuli presented in the periphery were investigated while top-down attention was directed to the central fixation point. The first two experiments investigated the neural responses to orientation contrast (experiment 1) and to luminance contrast (experiment 2). The results indicate that saliency is potentially encoded in a distributed fashion in the visual system and that a feature-independent saliency map is calculated late in the processing hierarchy. The third experiment used natural scenes as stimuli. Consistent with the results of the other two experiments, graded saliency was identified in striate and extrastriate visual cortex, in particular in posterior intraparietal sulcus (pIPS), potentially reflecting a representation of feature-independent saliency. Additionally, using multivariate pattern classification, information about the most salient positions could be decoded in more anterior brain regions, namely in anterior intraparietal sulcus (aIPS) and frontal eye fields (FEF). Taken together, the results suggest a distributed saliency processing of different low-level features in striate and extrastriate cortex that is potentially integrated to a feature-independent saliency representation in pIPS. Shifts of attention to the most salient positions are then prepared in aIPS and FEF. As participants were engaged in a fixation task, the saliency is presumably processed in an automatic manner.

Zusammenfassung

Permanent steht dem visuellen System des Menschen eine riesige Menge an Informationen zur Verfügung. Um aus dieser Informationsflut relevante Aspekte hervorzuheben, wird ein Filter benötigt. Aufmerksamkeit auf visuelle Reize kann dabei durch *top-down* Such-Strategien, z.B. wenn wir nach etwas Grünem suchen, oder durch *bottom-up* Eigenschaften des visuellen Reizes gesteuert werden. Die Eigenschaft einer bestimmten Position, aus einer visuellen Szene heraus zu stechen, wird als Salienz bezeichnet. Es wird angenommen, dass auf neuronaler Ebene eine Salienzkarte existiert, d.h. eine Karte, die die Salienz für jede Position des Gesichtsfeldes kodiert. Allerdings ist bis heute strittig, wo die Repräsentation einer solchen Karte im Gehirn lokalisiert sein könnte.

Im Rahmen dieser Dissertation wurden drei Experimente durchgeführt, die verschiedene Aspekte von *bottom-up* Salienz-Verarbeitung im menschlichen Gehirn mit Hilfe der funktionellen Magnetresonanztomographie (fMRT) untersuchten. Während die Aufmerksamkeit der Probanden auf einen Fixationspunkt gerichtet war, wurde die neuronale Reaktion auf unterschiedlich saliente Stimuli in der Peripherie untersucht. In den ersten zwei Experimenten wurde die neuronale Antwort auf Orientierungskontrast (Experiment 1) und Luminanzkontrast (Experiment 2) untersucht. Die Ergebnisse deuten darauf hin, dass Salienz möglicherweise verteilt im visuellen System kodiert ist und dass eine merkmalsunabhängige Salienzkarte relativ spät in der Verarbeitungshierarchie berechnet wird. Im dritten Experiment wurden natürliche Szenen als Stimuli verwendet. Im Einklang mit den Ergebnissen der ersten beiden Experimente wurde hier graduierte Salienz in frühen und späten visuellen Arealen identifiziert, insbesondere auch im posterioren intraparietalen Sulcus (pIPS), was auf eine Repräsentation merkmalsunabhängiger Salienz hindeuten könnte. Darüber hinaus konnten mit multivariater Mustererkennung, Informationen über die salientesten Positionen aus weiter anterior liegenden Hirnregionen, wie dem anterioren intraparietalen Sulcus (aIPS) und dem frontalen Augenfeld (FAF), dekodiert werden. Zusammengenommen deuten die Ergebnisse auf eine verteilte Salienzverarbeitung von unterschiedlichen low-level Merkmalen in frühen und späten visuellen Arealen hin, die möglicherweise zu einer merkmalsunabhängigen Salienzrepräsentation im pIPS zusammengefasst werden. Verschiebungen der Aufmerksamkeit zu den salientesten Positionen werden dann im aIPS und im FAF vorbereitet. Da die Probanden mit einer Fixationsaufgabe beschäftigt waren, wird die Salienz vermutlich automatisch verarbeitet.

List of Original Publications

This dissertation is based on the following research articles:

Study I

Bogler C, Bode S, Haynes JD. (2013). Orientation pop-out processing in human visual cortex. *Neuroimage,* 81: 73-80.

Study II

Betz T, Wilming N, Bogler C, Haynes JD, König P. (Submitted to Journal of Vision). Dissociation between saliency signals and activity in early visual cortex.

Study III

Bogler C, Bode S, Haynes JD. (2011). Decoding successive computational stages of saliency processing. *Curr Biol*, 21(19): 1667-71.

ABBREVIATIONS

aIPS	Anterior Intraparietal Sulcus
BOLD	Blood Oxygen-Level Dependent
dHb	Deoxyhaemoglobin
FEF	Frontal Eye Fields
fMRI	Functional Magnetic Resonance Imaging
GLM	General Linear Model
Hb	Oxyhaemoglobin
HR	Hemodynamic Response
ICA	Independent Component Analysis
IOR	Inhibition of Return
LFP	Local Field Potentials
MUA	Multi-Unit Activity
MVPA	Multivariate Pattern Analysis
NMR	Nuclear Magnetic Resonance
PCA	Principal Component analysis
pIPS	Posterior Intraparietal Sulcus
RFE	Recursive Feature Elimination
ROI	Region of Interest
V1	Primary Visual Cortex
WTA	Winner-Take-All

TABLE OF CONTENTS

1 INTRODUCTION

1.1 General Overview

Humans and animals are constantly confronted with the task to process a vast amount of information through the visual system. The amount of information is too high for every detail to be processed, so that filtering is needed. In Figure 1, a visual scene from a supermarket is shown. There are lots of red tomatoes but also two different fruits among them. The green lemon is easy to spot among all the other red items. However, the red apple in the lower right part of the image is somewhat harder to find, as this item has a similar color like the tomatoes. The visual search literature distinguishes between feature search and conjunction search (Treisman and Gelade, 1980). In feature search, targets are defined by one single dimension or feature, for example a green target among red distractors. Visual search for such targets is very efficient, fast, and the target pops-out. Furthermore, search times do not increase if the number of distractors increases. In conjunction search, targets are defined by a combination of two or more different features, each of which alone is also shared with the distractors, for example a green horizontal target bar with green vertical and red horizontal distractor bars. Search times for such displays are relatively slow and increase with the number of distractors. Models have been proposed to explain these phenomena in visual search. Both, the feature integration theory (Treisman and Gelade, 1980) and the guided search model (Wolfe et al., 1989; Wolfe, 1994) propose the existence of different feature maps, for example for orientation and color. A feature map is a spatial representation of the visual field coding the presence of one single feature. According to the feature integration theory and the guided search model, feature search is so efficient because only one of these feature maps is necessary to identify the position of the target. In conjunction search two or more feature maps have to be combined to find the position of the target. Classically it was assumed that all

potential target positions are scanned consecutively on the feature maps until the target is found or until every possible position is checked. Therefore search times in conjunction search increase linearly, and search in target absent displays is slower (Treisman and Gelade, 1980). However, there is evidence that the feature maps are combined in a parallel fashion and search times in conjunction search increase faster than expected by a strict linear model (Wolfe et al., 1989; Wolfe, 1994).

Because of their relationship to other elements present in the visual field, each object has its own visual saliency. Visual saliency is the property of objects in the visual field to automatically attract one's attention, or in VanRullen's (2003, p. 366) words: "... [saliency] is whatever renders visual objects or locations interesting to our visual system". A pop-out target in feature search is a prime example of a very salient object. The visual system is tuned to preferentially process salient locations. Salient locations are defined by stimulus properties alone (i.e. bottom-up). Bottom-up visual attention is the mechanism to preferentially process salient locations of the visual field and a built-in filter into the visual system to ensure that potentially interesting locations are processed as fast as possible. It has been shown that interesting objects are visually salient based on low-level visual features (Elazary and Itti, 2008). Therefore, it is efficient to attend to salient positions as it increases the chance to attend to interesting objects in a visual scene. In other words, the visual system evolved such that the properties that render a location salient to the visual system are the same properties that interesting objects create when they are placed into a scene. Saliency seems also to be involved in conscious decision processes when subjects are asked to identify interesting positions in pictures (Masciocchi et al., 2009). However, to date it is still a matter of debate how saliency computation is exactly implemented in the brain.

For this thesis I used functional magnetic resonance imaging (fMRI) to investigate how saliency is represented in the human brain. In the following chapters I will first introduce computational saliency models (Paragraph 1.2) and then review the existing literature about

saliency representation in the brain (Paragraph 1.3); finally I will give an overview over the applied methods (Chapter 2) and describe the conducted experiments (Chapter 3).

Figure 1: At a supermarket two different fruits were placed among the tomatoes. A salient green lemon in the top left part of the image automatically attracts one's attention. On the other hand, the red apple in the lower right part of the image, which has a similar color like the tomatoes, is harder to find.

1.2 Saliency Map Models

Computational saliency models describe how different visual features are processed to calculate a potential saliency representation. Most saliency models assume a representation of saliency on a spatial map that encodes the saliency for each position of the visual field, the so-called saliency map (Itti and Koch, 2001). Such computational models are interesting for several reasons. First, such models can improve our understanding of how saliency is calculated in the brain and, second, the models could be used in real world applications. For

example, saliency map models are used to evaluate the design of websites or, more specifically, the good, i.e. salient, placement of content or advertisement (see http://whitematter.de/ as an example). In computer vision, object identification algorithms could be applied to salient locations first, as it is known that at salient locations chances to find interesting objects are high (Elazary and Itti, 2008). This biologically inspired computation would reduce computational costs and also lead to more natural behaving robots, which explore salient locations first (Siagian and Itti, 2009). Another application of saliency models is image, and especially video, compression. Lossy compression algorithms could be applied with low compression rates at salient positions and with high compression rates, i.e. higher information loss, at non-salient positions to which attention will be rarely drawn (Itti, 2004; Li et al., 2011).

One of the best-known saliency models is the saliency map model from Itti and Koch (1998, 2000, 2001), which is an extension of the original model from Koch and Ullman (1985). The model is biologically inspired and incorporates many different processing stages, which makes it rather complex compared to other saliency models (see below). In this model, firstly, different low-level features are extracted from an input image (luminance, color, orientation, and motion) using linear filtering and differently oriented Gabor pyramids (Itti et al., 1998). For each feature, the local contrast is calculated using center-surround differences modeled by two-dimensional difference-of-Gaussians at different spatial scales (Itti and Koch, 2000). These feature maps are very similar to the feature maps of the feature integration model (Treisman and Gelade, 1980) or the guided search model (Wolfe et al., 1989). Finally, the feature maps are combined to the feature-independent saliency map. A winner-take-all (WTA) mechanism is thought to operate on the saliency map to select the position with the highest saliency. This is the position to which attention would be overtly or covertly directed. So far, the pure bottom-up saliency computation is described. Importantly, for the realistic modeling of the direction of attention over time, further processes have to be taken into

account. Top-down influences due to search strategies (for example spatial and feature attention) can interact with the bottom-up saliency computation (Baluch and Itti, 2011). Also, inhibition of return (IOR) (Klein, 2000), an inhibitory aftereffect to redirect attention to an already attended position, should be incorporated into a complete model. It is essential that IOR can influence (inhibit) saliency of already inspected locations on the saliency map, because this allows to model scan paths for free viewing of visual scenes. The saliency map would otherwise remain static, resulting in a model that always predicts that attention is directed to the same spatial position.

Itti and Koch's model has successfully been used to replicate psychophysiological results of visual search and saccade positions for free viewing (Itti and Koch, 2000; Parkhurst et al., 2002; Peters et al., 2005). Furthermore, the model can be used to design computer vision algorithms for robotics and video compression (Itti, 2004; Walther and Koch, 2006).

The V1 saliency map theory, an alternative computational saliency map model directly linked to a specific brain structure, assumes that saliency is calculated and exhaustively represented in primary visual cortex (Li, 2002; Zhaoping and May, 2007; Zhang et al., 2012) (please note that Li and Zhaoping is the same author, who published under different names). Zhaoping's model is inspired by the properties of V1 neurons. However, it is important to note that the architecture of the feature maps in Itti and Koch's model also resemble basic properties of neurons in V1, and most importantly, also those of other visual areas. In consequence, instead of accounting for visual saliency in general, the V1 saliency map model could be considered a reduced version of Itti and Koch's model that calculates saliency only for the features explicitly processed in primary visual cortex and, thereby, only at the spatial scale that corresponds to the receptive field size of V1 neurons. Furthermore, an explicit WTA mechanism does not exist in the V1 saliency map model, although Koch and Ullman (1985) describe a neural implementation of WTA. Thus, it might not be sufficient that the most

salient position is implicitly encoded in the saliency map. These restrictions cast serious doubts on the V1 model for visual saliency in its present form.

Finally, an alternative data-driven approach to calculate a saliency map was used by Kienzle et al. (2009). They recorded eye movements during the presentation of natural scenes. The most predictive features for saccades could be calculated from the images in combination with the recorded eye positions. Patches, i.e. features of an image that attract fixations, had a high contrast and a "corner-like" structure. The convolution of the most predictive feature patches with a new image is an estimation of the saliency map. This is an elegant and computationally efficient way for calculating a saliency map. Furthermore, it is possible to link the shape and size of the most predictive features to known properties of visual areas. Kienzle and colleagues concluded that the receptive fields in monkey superior colliculus matched best with their results. Their approach is similar to the identification of visual primitives, like gabor filters, that are used to describe processing in striate visual cortex as a filter bank (Jones et al., 1987; Jones and Palmer, 1987b; Jones and Palmer, 1987a). Although this outlook seems to become a promising approach, for the rest of this thesis only biologically inspired saliency map models like Itti and Koch's model or the V1 saliency map model will be considered.

1.3 Representation of Saliency in the Brain

Several groups have investigated the representation of saliency or saliency maps in the brain. Saliency information was found to be represented in several parts of the visual system, including the superior colliculus (Kustov and Robinson, 1996), the pulvinar (Robinson and Petersen, 1992; Shipp, 2004), V1 (Kastner et al., 1997; Nothdurft et al., 1999; Li, 1999; Li, 2002; Zhang et al., 2012), visual area V4 (Mazer and Gallant, 2003; Ogawa and Komatsu, 2006; Burrows and Moore, 2009), the parietal cortex (Gottlieb et al., 1998; Serences et al., 2005; Serences and Yantis, 2007; Geng and Mangun, 2009; Bisley and Goldberg, 2010) and the frontal eye fields (Thompson et al., 1997; Thompson and Bichot, 2005; Serences and

Yantis, 2007). The diversity of findings could be related to the various kinds of stimuli that were used. However, the term saliency was sometimes also used for both task-dependent, i.e. top-down, and stimulus-driven, i.e. bottom-up, processes and therefore some of the findings might be explained by this inconsistency. Thus, I will concentrate on bottom-up saliency in this thesis. In the next paragraph I will describe a saliency paradigm designed for fMRI experiments that minimizes effects of top-down attention.

1.4 The fMRI Saliency Paradigm

For the investigation of saliency on the behavioral level, Braun (1994) proposed a dual task paradigm: One task performed on stimuli shown at the center of a display was used for top-down attention control, and a second task performed on a salient stimulus in the periphery was used to measure the effect of saliency. The stimulus in the periphery was considered salient if properties of that stimulus could be consciously reported without cost for the main fixation task. This seems to be a promising approach, because it offers a measure of saliency (i.e. performance on the salient stimuli) and additionally controls for top-down attention (i.e. performance on the center of the display); however, in light of new findings a number of problems occur with this approach.

1. It has been shown that saliency or pop-out stimuli can be processed without conscious awareness (Lin et al., 2009; Hsieh et al., 2011). An unaware pop-out stimulus primes the processing at the spatial location where it was presented (Posner et al., 1980) if top-down attention is available (Hsieh et al., 2011).

2. It has also been shown that the parietal cortex is involved in bottom-up and top-down attention (Corbetta and Shulman, 2002; Serences and Yantis, 2007; Geng and Mangun, 2009; Bisley and Goldberg, 2010; Shomstein, 2012), although it responds more strongly to bottom-up attention (Geng and Mangun, 2009). Furthermore, brain regions that are involved in overt and covert shifts of attention are highly overlapping (Corbetta et al., 1998; Beauchamp et al.,

2001) and include the parietal cortex and the frontal eye fields. Therefore, we need a task that controls for top-down attention, and furthermore, we need to avoid any shifts of attention to salient stimuli at different positions. As a consequence, potential activations in the parietal cortex can be clearly attributed to the saliency of the stimuli and are not potentially related to shifts of attention to the salient stimuli. Fixation control with eye-tracking, which is difficult in fMRI experiments, is no sufficient control because covert shifts of attention can be correlated with activity in parietal cortex.

3. Top-down attention causes changes of activity in the visual cortex (Brefczynski and DeYoe, 1999; Gandhi et al., 1999; Kastner et al., 1999; Kastner and Ungerleider, 2000; Liu et al., 2007; Serences and Boynton, 2007). This is another reason to avoid (overt and covert) shifts of attention to the salient stimuli. Even if the top-down effects will occur later than the stimulus driven responses in the visual cortex, the sluggishness of the fMRI signal (see chapter 2.1) prevents a differentiation between stimulus driven and potential top-down effects within the same brain region.

In order to circumvent the problems discussed above and to bind top-down attention to the center of the screen, participants in all three experiments presented in this thesis were engaged in a simple but attentionally demanding task at the fixation point. The task continued during the whole course of the experiment. While subjects were solving the fixation task, we presented visual stimuli in the periphery. There was no task on these stimuli, and subjects were instructed that the stimulation could be ignored and that they should concentrate on the fixation task only. As a consequence neural responses in any brain region are very likely related to bottom-up saliency only and not to executed overt or covert shifts of attention or even spatial top-down attention to the periphery.

2 METHODS

In the three experiments discussed in the present thesis, we measured brain responses using fMRI. This chapter briefly describes the principles of fMRI (Paragraph 2.1). Subsequently, conventional mass univariate analyses (Paragraph 2.2) as well as more advanced and sensitive multivariate analytical approaches (Paragraph 2.3) are described since both types of analyses have been used to evaluate the acquired data.

2.1 Functional Magnetic Resonance Imaging (fMRI)

Here, I provide a very basic overview of the principles of fMRI. In-depth details can be found for example in Huettel and colleagues (2009).

MRI makes use of the fact that a high proportion of the brain consists of water. Atomic nuclei (for example H+ protons of water) have a spin which aligns parallel or antiparallel to an external magnetic field. Due to the fact that more spins align parallel than antiparallel, a net magnetization along the magnetic field axis is established (longitudinal magnetization). The aligned spins precess with the Lamor-frequency (ω), which is the product of the external static magnetic field strength (B_0) and the gyromagnetic ratio (γ), and which is specific for the type of atomic nucleus:

$$\omega = B_0 \gamma$$

Then, a radio frequency pulse (RF pulse) with frequency ω is sent from a transmitter coil, causing 1) the spins to flip and 2) to synchronize the precession of the spins. This results in a rotating net magnetization vector transversal to the external field that can be recorded by a detector coil. Nuclear magnetic resonance (NMR) imaging (Bloch, 1946; Bloch, 1953) as well as all modern MRI scanners are based on the described principle.

After the RF pulse the atomic nuclei will start to align with the external magnetic field again. This process of longitudinal relaxation is called spin-lattice relaxation and described with the time constant T_1. Additionally, the spins will also start to precess out of phase again due to

interactions with other spins in the neighborhood. This process is called spin-spin relaxation and described with the time constant T_2. Moreover other local field inhomogeneities cause a faster transversal relaxation than described by T_2 alone. This process is described with the time constant T_2^*. The relaxation times differ depending on the tissue properties, including concentration differences between oxygenated and deoxygenated hemoglobin. MRI and fMRI makes use of these relaxation times to image tissue differences.

To image the brain, the three dimensional coding of a volumetric picture element (voxel) requires the use of three additional magnetic gradients. The first field gradient is applied parallel to B_0 and causes the spins to precess at different frequencies. Single slices orthogonal to B_0 can then be selected with an RF impulse that only excites spins with a specific frequency. Within such a slice two other magnetic gradients are implemented to code for the exact spatial position. The spatial position is frequency coded in one direction and phase coded in the other direction. With an inverse Fourier transform the space-frequency coded 2-dimensional slice (the k-space) can be reconstructed.

Ogawa and colleagues (Ogawa et al., 1990a; Ogawa et al., 1990b; Ogawa et al., 1992) showed that MRI can be used to image the blood-oxygen-level dependent (BOLD) response. This is possible because deoxyhaemoglobin (dHb) and oxyhaemoglobin (Hb) have different magnetic properties, i.e. they are paramagnetic and diamagnetic, respectively. The paramagnetic dHb distorts the magnetic field and as a consequence the signal is decreased.

Neural activity causes alterations in the blood oxygenation because of an increase in local energy consumption. During glucose metabolism oxygen is extracted from the blood. This oxygen extraction is thought to be compensated with a local oversupply in oxygenated blood (the blood flow increases) and therefore an increased BOLD signal (caused by the washout of deoxygendated haemoglobin) (Villringer and Dirnagl, 1995). After the initial dip (Hu and Yacoub, 2012), which potentially reflects the relative increase of dHb due to the consumption of oxygenated haemoglobin, a peak in the BOLD response with a latency of 3-8 s after the

neural activity can be observed that reflects the oversupply in Hb. Neural activity is accompanied by this typical BOLD response, also called the haemodynamic response (HR).

It has been shown that local field potentials (LFP) with frequency bands between 30 and 150 Hz have a greater contribution to the BOLD response compared to multi-unit activity (MUA) (Logothetis et al., 2001; Logothetis, 2002; Logothetis, 2003; Logothetis and Pfeuffer, 2004; Logothetis and Wandell, 2004). However, the LFP reflects incoming perisynaptic activity. On the other hand, many of the neural interconnections are also feed-back or interconnections between neighboring neurons within the same brain region (Logothetis, 2008), so there is no clear answer to whether the BOLD response measured in one brain region reflects incoming signals from other brain regions and/or processing within that region.

During an fMRI experiment three-dimensional echo-planar images (Poustchi-Amin et al., 2001) covering the whole brain are typically acquired every 2 to 3 s. The time course of the measured response is then subject to different preprocessing steps such as motion correction, slice time correction and spatial smoothing before the statistical analyses can be conducted.

2.2 Univariate Analyses

The BOLD signal increases linearly with the contrast or the duration of a visual stimulation (Boynton et al., 1996). It is assumed that the relationship between neural activity and the BOLD signal is also linear in other brain regions. Overlapping of haemodynamic responses prevent a straightforward estimation of the local neuronal activity. Therefore, typically a general linear model (GLM), capable of modeling the overlaps, is applied to analyze fMRI time series in every single recorded voxel. Regressors for the different experimental conditions are constructed by convolving the onset vectors for each condition with the canonical HRF-function. The resulting design matrix (X) is then fitted to the recorded data (Y) with a GLM that minimizes the error (ε):

$$Y = X \beta + \varepsilon$$

Statistics on the single subject level are calculated based on how much variance the beta parameters (β) account for. Random effects analysis across subjects is then performed on the estimated beta parameters. Typically, the beta parameters, which reflect the activity for one condition in one voxel, are compared against the beta parameters that reflect the activity for another condition using a t-test (Friston et al., 2007). The analysis is performed separately for each voxel, therefore the statistical results have to be corrected for multiple comparisons (Nichols, 2012). This mass-univariate approach assumes that information about different conditions is represented in mean differences of the BOLD response in single voxels and furthermore that neighboring voxels show similar BOLD responses, as for example spatial smoothing is used to increase the signal to noise ratio for the statistical analysis.

2.3 Multivariate Analyses

The application of multivariate pattern analyses (MVPA) is relatively new to fMRI research (Haynes and Rees, 2006; Kriegeskorte et al., 2006). The rationale of the approach is that information about different conditions is represented in patterns across local groups of voxels, and that averaging across the voxels that would contribute to these patterns, as it is typically done in univariate analysis, will degrade the available information. Therefore, MVPA that considers the pattern information of two or more recorded voxels is potentially more sensitive compared to univariate analyses (Haynes and Rees, 2005). In this chapter I will briefly explain the necessary steps for conducting an MVPA. More detailed introductory tutorials for the application of MVPA to neuroimaging data are available (Formisano et al., 2008; Mur et al., 2009; Pereira et al., 2009; Lemm et al., 2011; Mahmoudi et al., 2012).

Typically, in MVPA data from a subset of the recorded data are used to train a model (see Paragraph 2.3.2). The experimental condition, or the "label", of new data that was not used during the training phase is then predicted by this model. The label of the new data can be

either *classified,* as in the case of two or more experimental conditions, or *regressed,* as in the case of a continuous variable (for example reaction times).

There are 3 main steps that are necessary for performing a MVPA:

2.3.1 Feature Selection

The first step of the analysis concerns feature selection. Feature selection raises two questions. The first is *what* the features are, and this is related to the preprocessing of the fMRI data. For example the raw fMRI data could be used for the analysis. Alternatively, less noisy estimates for the conditions could be used, based on the averaged response or on parameter estimates of a GLM (Mourão-Miranda et al., 2006). The next step of the feature selection concerns the question of from *where* in the brain we take the data. In typical whole brain fMRI experiments, activity in thousands of voxels is recorded, and usually the number of voxels exceeds the number of trials by several orders of magnitude. It is possible to perform MVPA on the whole brain; however, voxels that contain noise will have a relatively high impact and therefore a reliable separation between conditions might be impossible (Chu et al., 2012). A possible solution is selecting voxels that contain little noise, which will reduce the dimensionality of the data. Voxel selection based on univariate statistics was applied successfully (Haynes and Rees, 2005; Mourão-Miranda et al., 2006); however, the multivariate structure of the data is neglected for this type of feature selection ("… a variable that is completely useless by itself can provide a significant performance improvement when taken with others." Guyon and Elisseeff, 2003, p. 1165). Recursive feature elimination (RFE) (Guyon et al., 2002) reduces the number of features step-wise. RFE uses a nested (potentially also multivariate) method to rank the voxels according to their predictive value and stepwise excludes voxels that do contribute little to the prediction. Another possibility is dimensionality reduction by means of principal component analysis (PCA) (Mourão-Miranda et al., 2005; Brouwer and Heeger, 2009; Liu et al., 2012) or independent component analysis (ICA) (Anderson et al., 2010; Douglas et al., 2013) to calculate new features. These features

then explain a large part of the variance in the data but with a reduced dimensionality due to the combination of similar features to single components. Finally, region of interest (ROI) analyses for which the ROIs are defined based on separate functional localizer experiments (for example also retinotopic mapping, see Warnking et al., 2002; Wandell et al., 2007), or anatomically, are also possible.

A special type of ROI analysis is the so-called searchlight decoding approach (Kriegeskorte et al., 2006). With this method, a sphere with radius r is created around each of the N recorded voxels of the brain. MVPA is then performed N times, separately for all activity patterns contained in each searchlight cluster.

2.3.2 Prediction: Training and Testing

After the features are calculated and extracted, the data are split into two subsets: a training and a test dataset. The training data are used to train a multivariate model (classification: linear discriminant analysis, support vector classification, naïve Bayes, etc.; regression: multivariate regression, support vector regression, etc.). In the next step, the trained model is used to predict the labels of the test dataset. The accuracy of the prediction is assessed by comparing the predicted labels with the real labels of the data. For a better estimation of the real accuracy, usually the process of training and testing is repeated multiple times. For this, a leave-one-out cross-validation is often employed. In a leave-one-out cross-validation the data are divided into K subsets. A model is trained on all but one of these subsets. The left-out subset is later used for testing the model. This procedure is then repeated until each subset served as the test dataset once. The K accuracies, one for each of the cross-validation steps, are averaged. It is important to keep the training and test data in each cross-validation step strictly separate to avoid circularity (Kriegeskorte et al., 2009).

2.3.3. Statistics on the Prediction

The last analysis step is the group-level statistical test of the prediction accuracies. The consensus in the neuroimaging literature is to use a t-test to assess whether the mean accuracy is above the expected chance level (1/N conditions). Although for accuracies the requirements of a t-test are not perfectly satisfied (Stelzer et al., 2013), t-tests have the advantage that their computation is very efficient, for example using standard fMRI analysis packages (such as SPM: http://www.fil.ion.ucl.ac.uk/spm/) on the whole brain. In particular for whole brain searchlight decoding the calculation of t-tests in combination with multiple comparison correction methods is very useful. However, a more accurate way to assess the statistics of the accuracy can be achieved by comparing the real accuracy values with an empirical distribution obtained by permutation tests (Pereira and Botvinick, 2011). In a permutation test the training of the model and the prediction of new data points are repeated many thousand times with permuted labels. Permutation of the labels will destroy the correct data–label association, and in theory the prediction should be at chance level. Permutation tests for single subjects and for only a few brain regions/ROIs can be easily performed. However, permutation tests can become computationally demanding when they are performed for whole brain searchlight decoding with more than 30000 recorded voxels and for many subjects (Stelzer et al., 2013).

3 EXPERIMENTS

In this chapter three fMRI experiments that investigated the neural encoding of saliency signals in the near absence of top-down attention are summarized. The first experiment (Paragraph 3.1) investigated the processing of simple salient orientation pop-out stimuli in human visual cortex (Bogler et al., 2013). In the second experiment (Paragraph 3.2) we investigated the saliency of increased and decreased contrast in pink noise images using eye-tracking. We further used salient high- and low-contrast images and looked for contrast

independent saliency representations in the human brain (Betz et al., submitted). The third and last experiment (Paragraph 3.3) investigated the neural implementation of successive stages of a computational saliency model. In this experiment we used complex natural scenes as stimuli (Bogler et al., 2011).

3.1 Orientation Pop-Out Processing in Human Visual Cortex

In the first study we investigated the role of the visual cortex in the calculation of orientation contrast between a target stimulus and its surround. If a bar is surrounded by a set of homogenous bars, which have a different orientation than the target bar, then the latter pops out. Visual search for pop-out stimuli is fast, does not require attentional resources and runs in parallel for all items, therefore the search times are almost independent of the set size for pop-out stimuli (Treisman and Gelade, 1980).

A potential brain region encoding orientation pop-out is V1. Orientation selectivity in V1 has been shown in many studies with different methods (optical imaging: Ts'o et al., 1990; electrophysiological recordings: Hubel and Wiesel, 1962; fMRI: Tootell et al., 1998; Boynton and Finney, 2003). Furthermore, it has been shown that the response of neurons in V1 can be modulated by stimuli outside the classical receptive field (Knierim and van Essen, 1992; Sillito et al., 1995; Zipser et al., 1996; Li et al., 2000). Taken together, these findings suggest pop-out processing in V1. A direct representation of orientation pop-out in V1 was also reported in some studies (Kastner et al., 1997; Nothdurft et al., 1999; Zhang et al., 2012). However, Hegdé and Fellemann (2003) reported that responses of V1 neurons were modulated in a very similar way by stimuli outside the receptive field independent of whether these background stimuli were homogenous (pop-out condition) or heterogenous (non-pop-out condition). In other words, V1 could not discriminate between pop-out and non-pop-out stimuli in this study. Furthermore, Burrows and Moore (2009) reported that neurons in V4 showed the appropriate response profile for the representation of orientation pop-out. Finally,

a study in which area V4 was removed in monkeys (Schiller and Lee, 1991) demonstrated that search performance for pop-out stimuli was limited after ablation, suggesting an important role of V4 in pop-out processing. Taken together, there are studies suggesting a representation of pop-out in V1 while other studies suggest that V4 is more crucial in pop-out calculation. In our study we aimed to further investigate the role of V1, V4 and potentially other brain regions, in the calculation of orientation pop-out using whole-brain fMRI.

Psychophysical Experiment

As stimuli we used four different homogenous sets of distractor bars, which were all oriented 0°, 45°, 90° or 135° counterclockwise to the vertical direction (Figure 2A). First, we measured reaction times to pop-out stimuli in a behavioral experiment outside the scanner. In this experiment, one of two possible target bars (either on the left or right side of the screen) was rotated 15°, 30°, 45°, 60°, 75° or 90° counterclockwise *relative* to the distractor bars and served as the pop-out stimulus. The subjects' task was to detect the stimulus and indicate the side on which it appeared. It is important to note that pop-out is defined by the *relative* orientation between target and distractors. The absolute target orientation, which is encoded in V1 and could be decoded in fMRI experiments (Haynes and Rees, 2005; Kamitani and Tong, 2005), is not informative about the strength of the pop-out configuration. Subjects were faster in detecting the pop-out when the orientation contrast was higher. The accuracy in the performance was not different for the different orientation contrasts.

fMRI Experiment

Stimuli used in the fMRI experiment were very similar to the ones used in the behavioral experiment. The stimulus configurations had the same size and the same four distractor orientations were used. Only the orientation contrasts for the pop-out stimuli were slightly different with 0°, 30°, 60° or 90° relative to the distractor bars. Importantly, during stimulus exposure, subjects performed an attentionally demanding fixation task to control for top-down

attention (see above), rendering the orientation contrast stimuli in the background task-irrelevant. We used 16 conditions (4 orientation contrasts left X 4 orientation contrasts right).

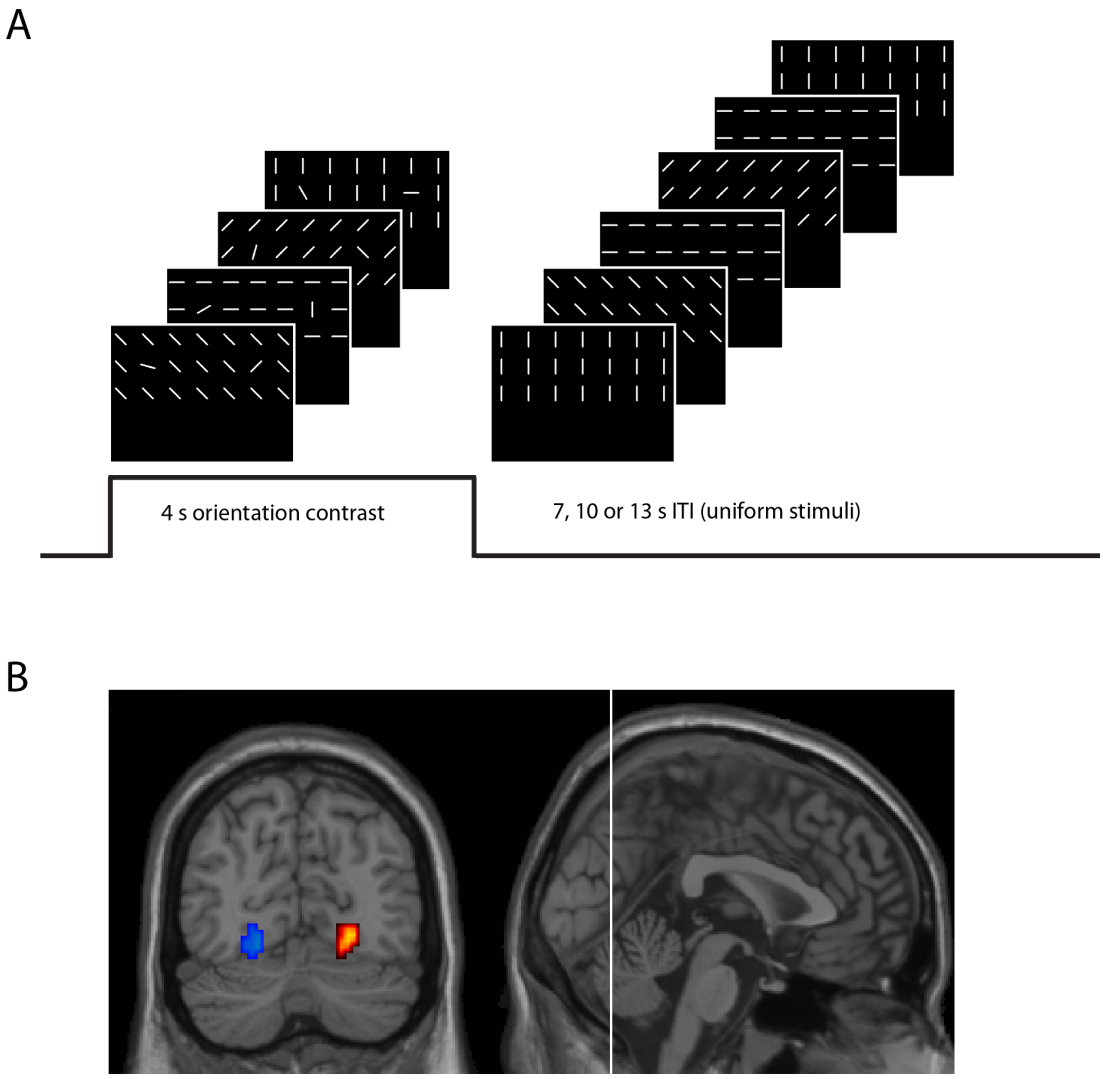

Figure 2: A: During the first 4 s of a trial one orientation contrast condition was presented. The orientation contrast between the target and the distractor bars was the same for the four stimuli although different background orientations were used. Homogenous stimuli were presented for 7, 10 or 13 s after the orientation contrast stimulation. B: BOLD response in V4 was correlated with the level of pop-out contralateral to the position of the pop-out.

During the experiment, we continuously presented either homogenous stimuli (that were used as the implicit baseline for the general linear model) or one of the 16 orientation contrast conditions (see Figure 2A). During one trial, the presentation of the orientation contrast stimulus was repeated 4 times and lasted 4 seconds. Importantly, all four possible combinations of distractor orientations and target stimulus orientations were used for each

orientation contrast condition; this allowed for extracting an estimate of contrast pop-out that was independent of absolute orientations. In the analyses we aimed to identify brain regions that showed a response profile that was compatible with the behavioral data from the psychophysical experiment. We performed a region of interest (ROI) analysis in anatomically defined ROIs in visual cortex and a whole brain analysis. In both analyses only the hemodynamic response in V4 was significantly modulated by orientation pop-out (see Figure 2B). The neural activity in V4 was higher for strong orientation contrasts with a similar non-linear response profile as observed for the reaction times.

The results confirm that the strength of orientation contrast modified the saliency of the stimuli. Stimuli with high orientation contrast were easier to detect, therefore the reaction times were faster. On the neural level, V4, and not V1, was shown to be involved in the calculation of orientation contrast. This demonstrates that, under specific circumstances, V1 can be blind to salient orientation contrast. This result is in conflict with the V1 saliency map theory (Li, 1999; Li, 2002).

3.2 Dissociation between Saliency Signals and Activity in Early Visual Cortex

In the second study we investigated the responses in visual cortex to stimuli with local contrast modifications. In most stimuli, saliency is correlated with luminance contrast, i.e. a salient position with a high contrast edge (see Study 1) also has increased luminance contrast at the same position. It is further known that responses in visual cortex are correlated with luminance contrast (Boynton et al., 1996; Goodyear and Menon, 1998). High contrast stimuli evoke higher activity in visual cortex compared to low contrast stimuli. Thus, activity in visual cortex does not necessarily code for saliency but could simply code for luminance contrast. We expected that regions that encode contrast-independent saliency would show increased activity for high and low contrast modifications compared to unmodified images. However, regions that encode luminance contrast instead of overall saliency were expected to

show different responses to the low compared to the high contrast conditions. The luminance contrast response could then potentially been taken into account for the calculation of saliency at a later processing stage.

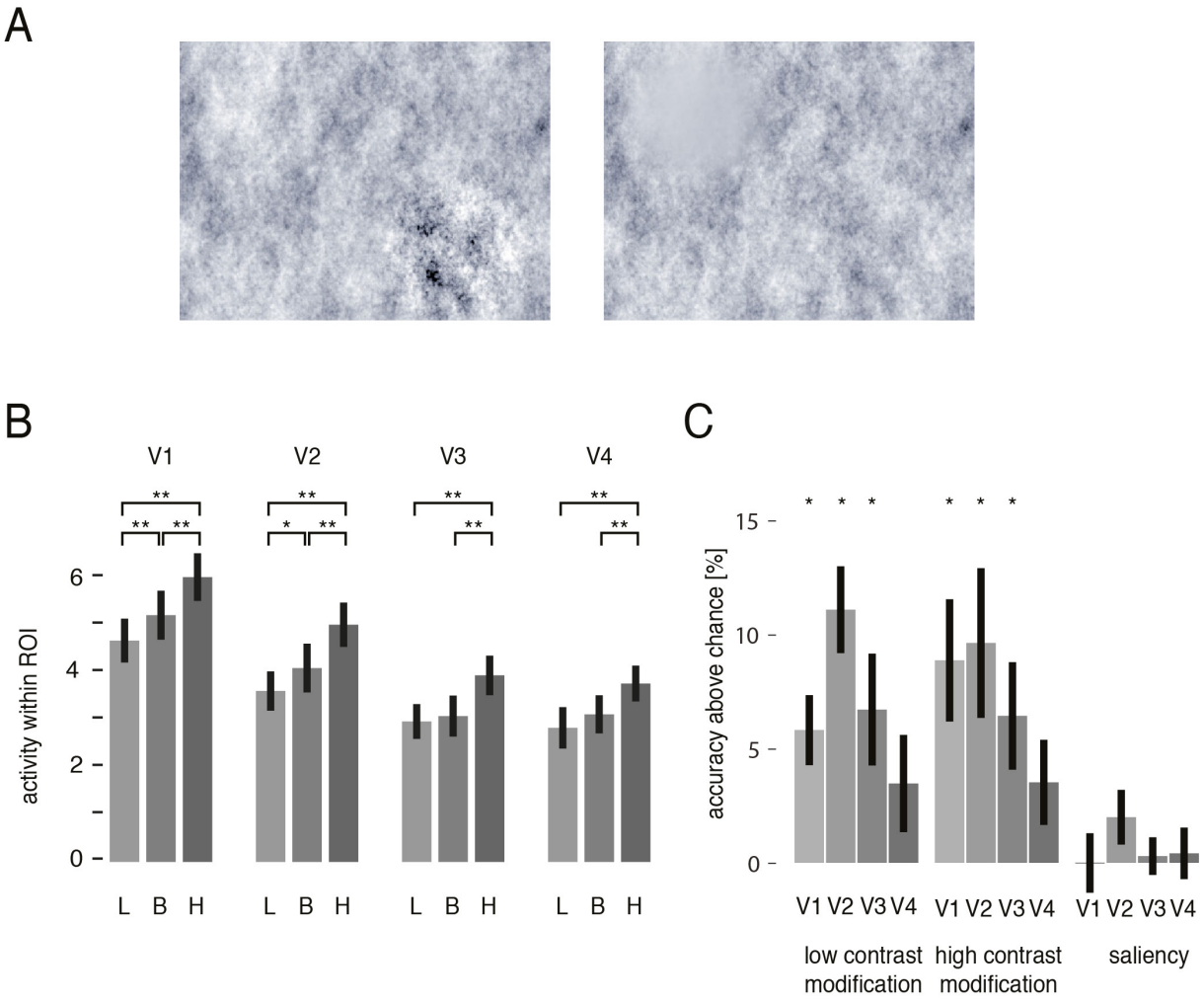

Figure 3: A) Pink noise stimuli with high contrast (left, lower right quadrant) and low contrast (right, top left quadrant) modifications. B) Averaged parameter estimates in different visual areas in the three contrast conditions (L = low, B = baseline, H = high) averaged across quadrants. In all areas, the high contrast condition evokes stronger BOLD signals compared to the baseline condition. Importantly, the low contrast condition evokes weaker BOLD signals compared to the baseline condition in V1 and V2 but never a stronger response. Error bars represent standard errors of the mean across subjects. Asterisks indicate significant differences between conditions (pairwise t-test, Holm-Bonferroni corrected; *: $p<0.05$, **: $p<0.01$). C) Decoding accuracies above chance level for the different visual areas. In V1, V2 and V3 it was possible to decode the contrast modification above chance level. In contrast, saliency could not be decoded. Asterisks indicate prediction performance significantly above chance level assessed by a t-test ($p<0.05$).

In this study, we used a set of pink noise stimuli, which have a power spectrum similar to the one of natural images (Einhäuser et al., 2006). The luminance contrast of the noise stimuli was then increased or decreased in one of the four quadrants of the stimuli (see Figure 3A).

In an eye tracking experiment we presented the stimuli and measured saccade positions while subjects performed a memory task. The aim of the memory task was to motivate the subjects to inspect the presented stimuli, so that they could answer potential questions about the stimuli after the presentation and after saccade positions were recorded. The results of the eye tracking experiment clearly demonstrated that decrements and increments of luminance contrast increase saliency. Thus, as both contrast modifications increase saliency, we could use these stimuli to disambiguate between the processing of luminance contrast and saliency by measuring BOLD response in an fMRI experiment.

In the fMRI experiment we presented the same pink noise stimuli with the contrast modifications while subjects performed an attentionally demanding fixation task to control for top-down attention (see above). The stimuli were completely irrelevant for the subjects' task. Region of interests (ROI) were defined for V1, V2, V3, and V4 by using retinotopic mapping procedures (Warnking et al., 2002; Wandell et al., 2007). For the main experiment, a GLM with 9 regressors (baseline; 4 x increased luminance contrast; 4 x decreased luminance contrast) was calculated, and parameter estimates were extracted from the ROIs.

We analyzed the mean activity in V1-V4 in the high contrast condition, the low contrast condition, and for the unmodified images. Furthermore, we used MVPA to analyze whether the activation patterns in the ROIs were informative about the contrast or the saliency modification.

Both univariate and multivariate ROI analyses could identify information about the luminance contrast in V1-V4. Importantly, contrast-independent saliency was not found to be represented in visual cortex.

Our result of contrast but not saliency representation in visual cortex (including V1) is in conflict with the V1 saliency hypothesis. We presume that an explicit representation of feature independent saliency is calculated later in the hierarchical processing stream and that the more specific information encoded in visual cortex is used at this later processing stage.

3.3 Decoding Successive Computational Stages of Saliency Processing

In the third study we aimed to identify neural correlates of different stages of bottom-up saliency processing. Itti and Koch's (2001) saliency map model assumes different computational stages (see Chapter 1.2 for a more detailed description of the model). In short, different low-level features (such as luminance, orientation, color, and motion) are extracted from an input image, and for each of these individual features different contrast maps are calculated. The corresponding contrast maps are then combined (i.e. integrated) into a saliency map that is, as a result, feature independent. As the next step in the processing hierarchy, the model suggests a non-linear transformation: The saliency map is assumed to be thresholded by a WTA mechanism, so that only the most salient position will be selected for a potential goal of overt or covert shifts of attention.

In this study, we were specifically interested in investigating the separate neural substrate of a) the representation of graded saliency as it would be coded in the saliency map and b) the WTA representation of only the most salient position. As stimuli we used photographs of natural scenes (Kienzle et al., 2009) as they combine different low-level features at multiple spatial scales. Furthermore, the visual system is tuned to natural stimuli (Einhäuser and König, 2010), therefore our stimuli were chosen to excite visual cortex in an optimal way.

Similar to the other fMRI experiments (see above), we presented our stimuli while subjects performed an attentionally demanding task on the fixation point during the whole course of the experiment to control for top-down attention (see above). We presented each of the 100 photographs for one second in each of the 5 runs of the experiment. During the one-second presentation the stimuli were switched on 3 times for 200 ms with pauses of 200 ms in between to increase signal to noise (Boynton et al., 2012).

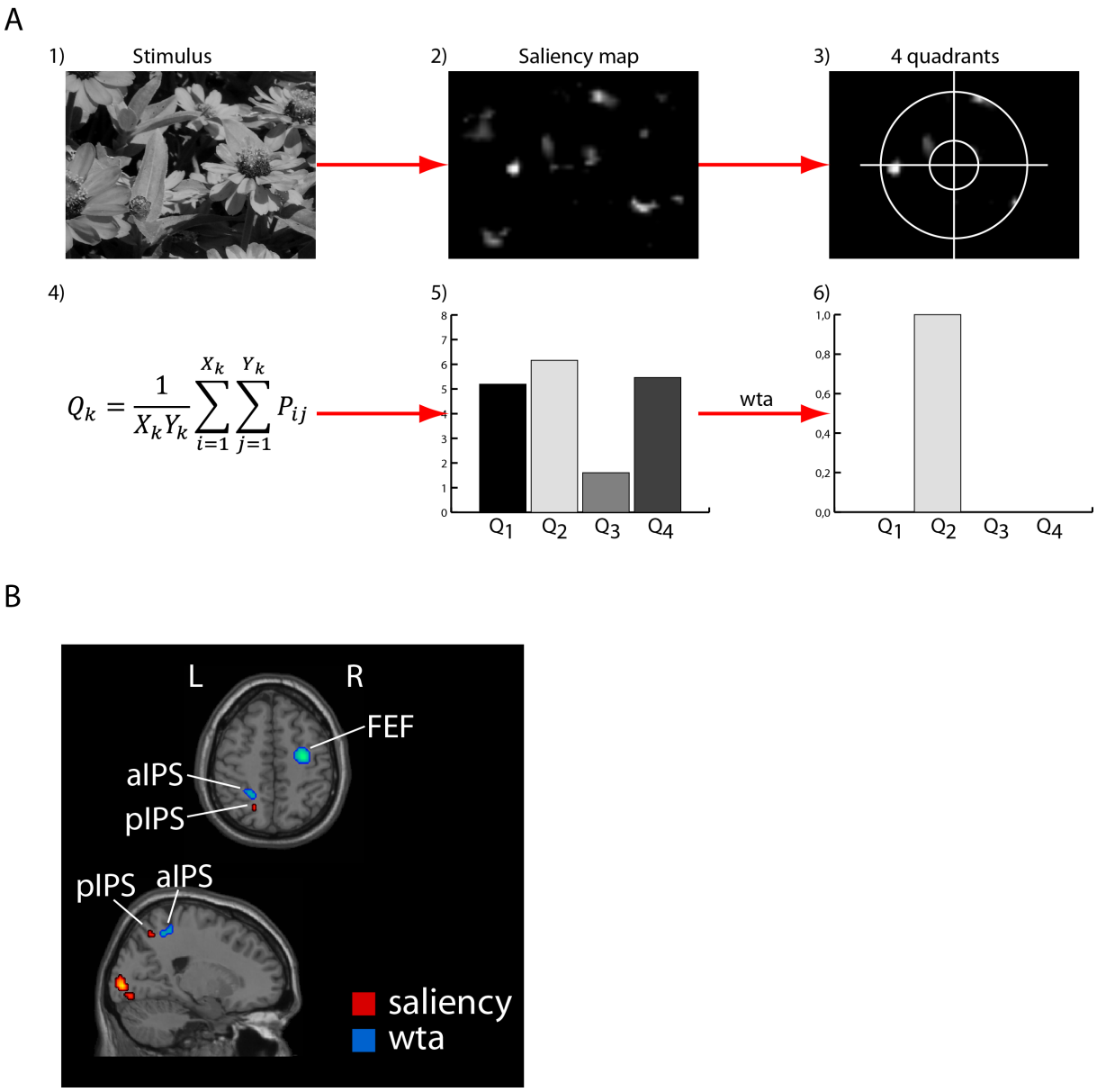

Figure 4: A) The data analysis was based on the predictions of a computational saliency map model. For each of the 100 different images of natural scenes (1) the corresponding saliency map (2) was calculated. The saliency was averaged across four quadrants (3-4), whereby central and peripheral regions of the visual field quadrants were not used. The average saliency for each quadrant (5) was used for the graded saliency analysis (Analysis I). A winner-take-all mechanism thresholded the graded saliency so that only the most salient quadrant remained (6). These were used for the WTA saliency analysis (Analysis II). B) Visual cortex and posterior intraparietal sulcus (pIPS) were correlated with graded saliency (red). The most salient quadrant could be decoded in the anterior IPS and frontal eye fields (FEF) (blue).

The data analysis was solely based on the predictions from Itti and Koch's saliency map model (see Figure 4A). For this, we calculated a saliency map for each of the 100 images with a Matlab based implementation of the saliency map model (Saliency Toolbox: http://www.saliencytoolbox.net/). The saliency maps were then divided into four quadrants and the graded saliency values within the quadrants were integrated. The centers of the

saliency maps were discarded since the participants of the experiment attended to the fixation task at the center of the screen. The four values that represented the mean graded saliency for the quadrants were used in the first fMRI analysis (*Analysis I: Graded saliency*), which was designed to find a representation of graded saliency. We assumed that the quadrant with the highest mean saliency value would be picked by a potential winner-take-all mechanism in the brain. Hence, in a next step we mirrored this process and defined, based on the computational saliency model, the most salient quadrant of the photographs. In the second fMRI analysis (*Analysis II: WTA saliency*) we then analyzed whether the most salient position (i.e. the most salient quadrant as derived from the model) was represented in brain activity, independent from the graded saliency map (as derived in Analysis I). Such a representation in the brain would most likely reflect the output of the potential winner-take-all mechanism that operates on the graded saliency map.

Analysis I: Graded saliency

We calculated a GLM with one regressor for the onsets of the visual stimulation and an additional parametric regressor for the graded saliency value. BOLD responses in striate and extrastriate visual cortex and in the posterior intraparietal sulcus (pIPS) were found to be significantly modulated by the graded saliency (see Figure 4B).

Analysis II: WTA saliency

For the second analysis, we calculated a GLM with onset regressors for each of the four conditions, representing trials in which each of the four quadrants was the most salient one, according to the model. However, conventional mass univariate analysis revealed no significant mean differences between the four conditions in any brain region. As the expected representation could be represented at a finer scale, more sensitive multivariate searchlight pattern classification analysis was used to decode the most salient condition from brain activity. For this analysis, we extracted parameter estimates of the GLM within a spherical searchlight from a subset of the data to train a support vector machine and in a next step

predicted the condition of parameter estimates that were not used in the training step. This was done repeatedly (for each individual subject) until each voxel served as the center of the searchlight once (see Chapter 2.3 for a detailed description of the method). Searchlight decoding revealed that information about the most salient quadrant was encoded in bilateral anterior intraparietal sulcus (aIPS) and bilateral frontal eye fields (FEF) (see Figure 4B).

The results suggest that different computational stages of a saliency model are represented in distinct brain regions. Graded saliency was found to be encoded in striate and extrastriate visual cortex and also in pIPS. The output of the WTA computation was encoded in more anterior brain regions, in aIPS and FEF. Furthermore, the results suggest that the calculation of saliency signals is performed automatically for unattended stimuli.

4 CONCLUSION AND FURTHER DIRECTIONS

The three studies that constitute the present thesis were aimed to investigate different aspects of visual bottom-up saliency processing in the human brain. In the first study we investigated how retinotopically organized brain regions in the visual cortex responded to salient changes in orientation contrast. In this study we identified a representation of orientation contrast in extrastriate cortex V4. Therefore, we could demonstrate that there are cases in which V1 is "blind" to orientation pop-out. The results are in conflict with the V1 saliency map hypothesis, which proposes the calculation and representation of saliency already at the level of V1. While saliency and orientation pop-out has been shown in V1 before (Kastner et al., 1997; Nothdurft et al., 1999; Li, 1999; Li, 2002; Zhang et al., 2012), none of the previous studies have investigated the relative orientation contrast. The discrepancy in findings could therefore result from the new feature of our design, which combines different absolute orientations into one condition. Alternatively, the null findings in V1 could also be related to the size of the bar stimuli used, which were relatively large, with 2.2°. Receptive field sizes of V1 neurons of 1.2° have been reported (Essen and Zeki, 1978; Snodderly and Gur, 1995;

Hegdé and Felleman, 2003), whereas receptive field sizes between 2.5° and 5° have been reported for V4 neurons (Burrows and Moore, 2009). Therefore, the bar stimuli could potentially be better processed by neurons of V4 rather than V1. In the context of Itti and Koch's saliency map model (Itti and Koch, 2001), orientation contrast stimuli are not suited to differentiate between the representations of a feature contrast map for orientation and a feature-independent saliency map because both maps are highly correlated in the case of using artificial orientation pop-out stimuli (see also below).

The second experiment was specifically designed to dissociate between brain responses that encode representations of luminance contrast and saliency. In an eye-tracking experiment we could demonstrate that both increments and decrements of luminance contrast are visually salient. Since it is known that responses in visual cortex are correlated with luminance contrast (Boynton et al., 1996; Goodyear and Menon, 1998), we used the luminance contrast stimuli to dissociate brain regions that encode luminance contrast and saliency. We showed that responses in V1, V2, V3, and V4 were correlated with luminance contrast only, and a representation of saliency, independent of luminance-contrast, could not be identified in visual cortex, not even with the application of more sensitive multivariate methods. The results are, again, not compatible with the V1 saliency map hypothesis because luminance contrast but not saliency was represented in V1. Notably, we identified a representation of luminance contrast also in V4. The results suggest that V4, which was identified in the first experiment as encoding orientation contrast, is not the region in which a representation of the saliency map is implemented. The luminance and orientation contrast, which were shown to be represented in visual cortex, are potentially used for the calculation of a feature independent saliency map at a later stage of the processing hierarchy.

In the third experiment we tested the assumption of a computational saliency map model. Specifically, we tested whether a graded saliency map representation and a WTA representation of the most salient location are encoded in distinct brain regions. Together with

the results of experiment 1 and 2, those of experiment 3 suggest that the representations of graded saliency or different feature maps are distributed throughout the visual system, including the striate and extrastriate cortex. Furthermore, the pIPS was found to be involved in the calculation of graded saliency. The most salient position, i.e. the position in the visual field where attention would be directed to, was encoded in more anterior brain regions, aIPS and FEF.

The results of all three experiments conflict with the V1 saliency map hypothesis (Li, 2002). In the first study we could not identify information about orientation contrast or saliency at the level of V1, as would have been predicted by the V1 saliency map hypothesis, but instead at the level of V4. In the second study, luminance contrast, but not saliency, was identified in V1, which again conflicts with the V1 saliency map hypothesis. Finally, the distributed representation of graded saliency in the striate and extrastriate cortex could be compatible with the V1 saliency map hypothesis if feed-forward processing from V1 to higher visual areas was assumed; however, WTA saliency was encoded in distinct brain regions in aIPS and FEF. While the V1 saliency map hypothesis proposes that WTA saliency is implicitly encoded in V1, our findings do not support this hypothesis.

In the three studies we used increasingly complex visual stimuli. While in the first study rather artificial stimuli (differently oriented bars) were used, the stimuli used in the second study shared the amplitude spectra of natural scenes (pink noise with a 1/f power spectral density) (Einhäuser et al., 2006). And finally, in the third study, we used photographs of natural scenes. The visual system has evolved to process natural stimuli with their inherent image statistics; therefore, despite the potential limits in experimental control, natural scenes offer unique advantages, in particular for the investigation of automatic processing (Einhäuser and König, 2010). Furthermore, natural scenes combine different low level features and therefore offer the possibility for investigating *feature independent* saliency. In the first and second study the saliency of a position in the visual field was dependent on only one single

feature, orientation or luminance contrast respectively. Importantly, in the second study the relationship between saliency and luminance contrast was not linear, which allowed us to dissociate responses to luminance contrast from responses to saliency. However, it would have not been sufficient to identify a brain region that responded with increased activity to high and low luminance contrast compared to baseline as a signature of *feature independent* saliency. Only the natural scenes that were used as stimuli in the third study combined different low-level features, and therefore the identified graded saliency representation could potentially be regarded as a representation of *feature independent* saliency. Based on a large number of previous studies that demonstrated responses to different visual stimuli in striate and extrastriate visual cortex and the results of the first and second experiments of this thesis, we propose that the graded saliency representation that was found in striate and extrastriate cortex in the third study is related to feature contrast maps. The representation of graded saliency in pIPS, which was also found in experiment 3, could potentially be a combination of different feature contrasts at this stage of the processing hierarchy. A WTA representation of the most salient quadrant in more anterior but not overlapping aIPS might further corroborate this hypothesis, as a WTA mechanism needs to operate on feature independent saliency or combine all feature maps. However, we could not directly test this hypothesis in our experiments because the individual feature maps are highly correlated (Parkhurst et al., 2002). Dominant feature dimensions in the stimuli potentially drive the overall saliency (e.g., the orientation contrast map being dominant for saliency when a single differentially oriented bar is presented in a uniform field of bars) (Parkhurst et al., 2002). For the stimuli in the third study we also found high correlations between the intensity and orientation contrast maps and the resulting saliency map. Furthermore, redundancy between the channels is likely to occur (Parkhurst et al., 2002), since the channels are not orthogonal but partly rely on the same information. For example, the border between a dark and a light area on an image resembles an edge. As a consequence, the position would be highlighted in both the orientation feature

contrast map and the luminance feature contrast map, and would therefore be coded as highly salient. Because of this high intercorrelation between the saliency map and the feature contrast maps it is not easy to distinguish between representations of the feature independent saliency map and the individual feature contrast maps. Therefore, further studies using carefully selected sets of visual stimuli are necessary. These studies need to combine different feature contrast maps in a way that the contribution of the individual features to the overall saliency can be distinguished and dissociated from each other.

The participants of all the three studies were engaged in similar attentionally demanding fixation tasks at the center of the screen. The task was designed to bind top-down attention to the center of the screen, while the stimuli of interest for the experiment were presented at the periphery. Previous studies suggest that top-down and bottom-up attention share similar neural networks (Corbetta and Shulman, 2002; Serences and Yantis, 2007; Geng and Mangun, 2009; Bisley and Goldberg, 2010; Shomstein, 2012). Therefore, we have to control for top-down attention to avoid that potential activation differences in a brain region are related to top-down instead of bottom-up attention. The distinction between bottom-up saliency and top-down attention was not always clear in previous studies that used the term saliency to describe behaviorally relevant stimuli (Corbetta and Shulman, 2002). In contrast, in the present three experiments, we focused on bottom-up saliency and made sure that the stimuli were completely irrelevant for the subjects. Subjects could ignore the stimuli while they solved the fixation task. Different results emerged in earlier studies could be due to the use of different definitions of saliency. Therefore, we propose that our results shed light on how automatic stimulus driven saliency is processed in the human brain.

Other authors have suggested the term “priority map” for a map where stimulus-driven and top-down effects are combined (Bisley and Goldberg, 2010; Fecteau and Munoz, 2006). This definition and the investigation of pure bottom-up, pure top-down, and the interaction of both,

is potentially more promising for understanding the neural architecture of attention processing in the brain.

In experiment 3 we demonstrated that WTA saliency is encoded in aIPS and FEF. These findings are inconsistent with the results of experiment 1 and 2 in which no representations of WTA could be identified in these brain regions. One explanation for this discrepancy could be that the stimuli in the three experiments were very different and natural scenes were used only in the third experiment. It has been shown that the response to pop-out stimuli in parietal cortex is reduced, when the stimuli were ignored (Ipata et al., 2006). Potentially the influence of bottom-up attention might be easier to suppress when the stimuli are easy to predict. This suppression might be reflected in weaker responses in more anterior brain regions. Within the first two experiments, very similar stimuli of either different oriented bars or pink noise with luminance contrast modifications were presented. In contrast, in experiment 3, 100 photographs of natural scenes were used; this made it harder to make predictions about the next stimulus. Please note that in the third experiment we conducted an additional control experiment outside the scanner to investigate whether the fixation task was demanding enough so that subjects could not consciously report the content or guess the saliency of the presented natural scenes. The results suggest that processing of natural scenes is harder to suppress and therefore processed at higher stages in the visual processing hierarchy, up to pIPS, aIPS, and FEF even though the stimuli are processed automatically. Importantly, conventional mass-univariate analysis approaches could not identify representations of WTA saliency in aIPS and FEF in the third study. Only more sensitive multivariate approaches that take into account the information of fine-grained activation patterns could identify the encoded WTA saliency in aIPS and FEF. This further suggests that the representation of bottom-up saliency in more anterior regions is relatively weak. Only the combination of strong and unpredictable visual stimuli with sensitive analyses approaches could demonstrate that bottom-up saliency is encoded in aIPS and even in frontal brain regions like FEF.

The results of the third study further suggest that saliency is represented in a more graded fashion in striate visual cortex, extrastriate visual cortex, and in pIPS, but in a WTA or categorical fashion in aIPS and FEF. Previous studies found that FEF and parts of the IPS control top-down attention (Corbetta and Shulman, 2002). Furthermore, it has been shown that activity in visual cortex can be modulated by top-down attention (Brefczynski and DeYoe, 1999; Gandhi et al., 1999; Kastner et al., 1999; Kastner and Ungerleider, 2000; Liu et al., 2007; Serences and Boynton, 2007) and by direct stimulation of FEF (Moore and Armstrong, 2003; Armstrong et al., 2006; Ruff et al., 2006). Taken together, these results suggest that top-down attention potentially originates in more frontal regions and modulates activity in visual cortex. The results of the third study suggest a processing hierarchy that proceeds in the opposite direction from early visual cortex to FEF. Importantly, in more anterior brain regions the encoded saliency information changes to a more WTA representation. This WTA representation is in line with the fact that spatial top-down attention is usually directed to only one single position and not distributed in a graded fashion through the whole visual field. Therefore, the distinction between graded vs. WTA saliency could extend the classical bottom-up vs. top-down perspective.

To summarize, we have conducted three fMRI experiments. The results suggest that bottom-up saliency calculation is implemented in a distributed fashion in striate and extrastriate visual cortex. Our results further suggest that a feature-independent saliency representation is calculated relatively late in the hierarchy of the visual system. Saliency of unattended stimuli can also be encoded in more anterior brain regions, like aIPS and FEF, which might prepare for shifts of attention. Therefore, the representation of saliency shifts from a more graded to a binary or categorical WTA representation as it reaches more anterior brain regions.

5 REFERENCES

Anderson, A., Dinov, I. D., Sherin, J. E., Quintana, J., Yuille, A. L., and Cohen, M. S. (2010). Classification of spatially unaligned fmri scans. *Neuroimage*, 49(3):2509–2519.

Armstrong, K. M., Fitzgerald, J. K., and Moore, T. (2006). Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron*, 50(5):791–798.

Baluch, F. and Itti, L. (2011). Mechanisms of top-down attention. *Trends Neurosci*, 34(4):210–224.

Beauchamp, M. S., Petit, L., Ellmore, T. M., Ingeholm, J., and Haxby, J. V. (2001). A parametric fmri study of overt and covert shifts of visuospatial attention. *Neuroimage*, 14(2):310–321.

Betz, T., Wilming, N., Bogler, C., Haynes, J.-D., and König, P. (submitted to Journal of Vision). Dissociation between saliency signals and activity in early visual cortex.

Bisley, J. W. and Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annu Rev Neurosci*, 33:1–21.

Bloch, F. (1946). Nuclear induction. *Phys. Rev.*, 70(7-8):460–474.

Bloch, F. (1953). The principle of nuclear induction. *Science*, 118(3068):425–430.

Bogler, C., Bode, S., and Haynes, J.-D. (2011). Decoding successive computational stages of saliency processing. *Curr Biol*, 21(19):1667–1671.

Bogler, C., Bode, S., and Haynes, J.-D. (2013). Orientation pop-out processing in human visual cortex. *Neuroimage*, 81C:73–80.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *J Neurosci*, 16(13):4207–4221.

Boynton, G. M., Engel, S. A., and Heeger, D. J. (2012). Linear systems analysis of the fmri signal. *Neuroimage*, 62(2):975–984.

Boynton, G. M. and Finney, E. M. (2003). Orientation-specific adaptation in human visual cortex. *J Neurosci*, 23(25):8781–8787.

Braun, J. (1994). Visual search among items of different salience: removal of visual attention mimics a lesion in extrastriate area v4. *J Neurosci*, 14(2):554–567.

Brefczynski, J. A. and DeYoe, E. A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nat Neurosci*, 2(4):370–374.

Brouwer, G. J. and Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *J Neurosci*, 29(44):13992–14003.

Burrows, B. E. and Moore, T. (2009). Influence and limitations of popout in the selection of salient visual stimuli by area v4 neurons. *J Neurosci*, 29(48):15169–15177.

Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., Lin, C., and for the Alzheimer's Disease Neuroimaging Initiative (2012). Does feature selection improve classification accuracy? impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage*, 60(1):59–70.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Petersen, S. E., Raichle, M. E., Essen, D. C. V., and Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron*, 21(4):761–773.

Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*, 3(3):201–215.

Douglas, P. K., Lau, E., Anderson, A., Head, A., Kerr, W., Wollner, M., Moyer, D., Li, W., Durnhofer, M., Bramen, J., and Cohen, M. S. (2013). Single trial decoding of belief decision making from eeg and fmri data using independent components features. *Front Hum Neurosci*, 7:392.

Einhäuser, W. and König, P. (2010). Getting real-sensory processing of natural stimuli. *Curr Opin Neurobiol*, 20(3):389–395.

Einhäuser, W., Rutishauser, U., Frady, E. P., Nadler, S., König, P., and Koch, C. (2006). The relation of phase noise and luminance contrast to overt attention in complex visual stimuli. *J Vis*, 6(11):1148–1158.

Elazary, L. and Itti, L. (2008). Interesting objects are visually salient. *J Vis*, 8(3):3.1–315.

Essen, D. C. and Zeki, S. M. (1978). The topographic organization of rhesus monkey prestriate cortex. *J Physiol*, 277:193–226.

Fecteau, J. H. and Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends Cogn Sci*, 10(8):382–390.

Formisano, E., Martino, F. D., and Valente, G. (2008). Multivariate analysis of fmri time series: classification and regression of brain responses using machine learning. *Magn Reson Imaging*, 26(7):921–934.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., and Penny, W. (2007). *Statistical parametric mapping the analysis of funtional brain images*. Elsevier/Academic Press, Amsterdam; Boston.

Gandhi, S. P., Heeger, D. J., and Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proc Natl Acad Sci U S A*, 96(6):3314–3319.

Geng, J. J. and Mangun, G. R. (2009). Anterior intraparietal sulcus is sensitive to bottom-up attention driven by stimulus salience. *J Cogn Neurosci*, 21(8):1584–1601.

Goodyear, B. G. and Menon, R. S. (1998). Effect of luminance contrast on bold fmri response in human primary visual areas. *J Neurophysiol*, 79(4):2204–2207.

Gottlieb, J. P., Kusunoki, M., and Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391(6666):481–484.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3):389–422.

Haynes, J.-D. and Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci*, 8(5):686–691.

Haynes, J.-D. and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat Rev Neurosci*, 7(7):523–534.

Hegdé, J. and Felleman, D. J. (2003). How selective are v1 cells for pop-out stimuli? *J Neurosci*, 23(31):9968–9980.

Hsieh, P.-J., Colas, J. T., and Kanwisher, N. (2011). Pop-out without awareness: unseen feature singletons capture attention only when top-down attention is available. *Psychol Sci*, 22(9):1220–1226.

Hu, X. and Yacoub, E. (2012). The story of the initial dip in fmri. *Neuroimage*, 62(2):1103–1108.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160:106–154.

Huettel, S., Song, A., and McCarthy, G. (2009). *Functional Magnetic Resonance Imaging*. Sinauer Associates, Incorporated.

Ipata, A. E., Gee, A. L., Gottlieb, J., Bisley, J. W., and Goldberg, M. E. (2006). Lip responses to a popout stimulus are reduced if it is overtly ignored. *Nat Neurosci*, 9(8):1071–1076.

Itti, L. (2004). Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Trans Image Process*, 13(10):1304–1318.

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res*, 40(10-12):1489–1506.

Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nat Rev Neurosci*, 2(3):194–203.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259. cited By (since 1996) 1147.

Jones, J. P. and Palmer, L. A. (1987a). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol*, 58(6):1233–1258.

Jones, J. P. and Palmer, L. A. (1987b). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J Neurophysiol*, 58(6):1187–1211.

Jones, J. P., Stepnoski, A., and Palmer, L. A. (1987). The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *J Neurophysiol*, 58(6):1212–1232.

Kamitani, Y. and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat Neurosci*, 8(5):679–685.

Kastner, S., Nothdurft, H. C., and Pigarev, I. N. (1997). Neuronal correlates of pop-out in cat striate cortex. *Vision Res*, 37(4):371–376.

Kastner, S., Pinsk, M. A., Weerd, P. D., Desimone, R., and Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, 22(4):751–761.

Kastner, S. and Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu Rev Neurosci*, 23:315–341.

Kienzle, W., Franz, M. O., Schölkopf, B., and Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *J. Vis.*, 9(5):1–15.

Klein (2000). Inhibition of return. *Trends Cogn Sci*, 4(4):138–147.

Knierim, J. J. and van Essen, D. C. (1992). Neuronal responses to static texture patterns in area v1 of the alert macaque monkey. *J Neurophysiol*, 67(4):961–980.

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol*, 4(4):219–227.

Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. *Proc Natl Acad Sci U S A*, 103(10):3863–3868.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., and Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci*, 12(5):535–540.

Kustov, A. A. and Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, 384(6604):74–77.

Lemm, S., Blankertz, B., Dickhaus, T., and Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *Neuroimage*, 56(2):387–399.

Li, W., Thier, P., and Wehrhahn, C. (2000). Contextual influence on orientation discrimination of humans and responses of neurons in v1 of alert monkeys. *J Neurophysiol*, 83(2):941–954.

Li, Z. (1999). Contextual influences in v1 as a basis for pop out and asymmetry in visual search. *Proc Natl Acad Sci U S A*, 96(18):10530–10535.

Li, Z. (2002). A saliency map in primary visual cortex. *Trends Cogn Sci*, 6(1):9–16.

Li, Z., Qin, S., and Itti, L. (2011). Visual attention guided bit allocation in video compression. *Image and Vision Computing*, 29(1):1 – 14.

Lin, J. Y., Murray, S. O., and Boynton, G. M. (2009). Capture of attention to threatening stimuli without perceptual awareness. *Curr Biol*, 19(13):1118–1122.

Liu, F., Guo, W., Yu, D., Gao, Q., Gao, K., Xue, Z., Du, H., Zhang, J., Tan, C., Liu, Z., Zhao, J., and Chen, H. (2012). Classification of different therapeutic responses of major depressive disorder with multivariate pattern analysis method based on structural mr scans. *PLoS One*, 7(7):e40968.

Liu, T., Larsson, J., and Carrasco, M. (2007). Feature-based attention modulates orientation-selective responses in human visual cortex. *Neuron*, 55(2):313–323.

Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos Trans R Soc Lond B Biol Sci*, 357(1424):1003–1037.

Logothetis, N. K. (2003). The underpinnings of the bold functional magnetic resonance imaging signal. *J Neurosci*, 23(10):3963–3971.

Logothetis, N. K. (2008). What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–878.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fmri signal. *Nature*, 412(6843):150–157.

Logothetis, N. K. and Pfeuffer, J. (2004). On the nature of the bold fmri contrast mechanism. *Magn Reson Imaging*, 22(10):1517–1531.

Logothetis, N. K. and Wandell, B. A. (2004). Interpreting the bold signal. *Annu Rev Physiol*, 66:735–769.

Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., and Brovelli, A. (2012). Multivoxel pattern analysis for fmri data: a review. *Comput Math Methods Med*, 2012:961257.

Masciocchi, C. M., Mihalas, S., Parkhurst, D., and Niebur, E. (2009). Everyone knows what is interesting: salient locations which should be fixated. *J Vis*, 9(11):25.1–2522.

Mazer, J. A. and Gallant, J. L. (2003). Goal-related activity in v4 during free viewing visual search. evidence for a ventral stream visual salience map. *Neuron*, 40(6):1241–1250.

Moore, T. and Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, 421(6921):370–373.

Mourão-Miranda, J., Bokde, A. L. W., Born, C., Hampel, H., and Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional mri data. *Neuroimage*, 28(4):980–995.

Mourão-Miranda, J., Reynaud, E., McGlone, F., Calvert, G., and Brammer, M. (2006). The impact of temporal compression and space selection on svm analysis of single-subject and multi-subject fmri data. *Neuroimage*, 33(4):1055–1065.

Mur, M., Bandettini, P. A., and Kriegeskorte, N. (2009). Revealing representational content with pattern-information fmri–an introductory guide. *Soc Cogn Affect Neurosci*, 4(1):101–109.

Nichols, T. E. (2012). Multiple testing corrections, nonparametric methods, and random field theory. *Neuroimage*, 62(2):811–815.

Nothdurft, H. C., Gallant, J. L., and Essen, D. C. V. (1999). Response modulation by texture surround in primate area v1: correlates of "popout" under anesthesia. *Vis Neurosci*, 16(1):15–34.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990a). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A*, 87(24):9868–9872.

Ogawa, S., Lee, T. M., Nayak, A. S., and Glynn, P. (1990b). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn Reson Med*, 14(1):68–78.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., and Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A*, 89(13):5951–5955.

Ogawa, T. and Komatsu, H. (2006). Neuronal dynamics of bottom-up and top-down processes in area v4 of macaque monkeys performing a visual search. *Exp Brain Res*, 173(1):1–13.

Parkhurst, D., Law, K., and Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Res*, 42(1):107–123.

Pereira, F. and Botvinick, M. (2011). Information mapping with pattern classifiers: a comparative study. *Neuroimage*, 56(2):476–496.

Pereira, F., Mitchell, T., and Botvinick, M. (2009). Machine learning classifiers and fmri: a tutorial overview. *Neuroimage*, 45(1 Suppl):S199–S209.

Peters, R. J., Iyer, A., Itti, L., and Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Res*, 45(18):2397–2416.

Posner, M. I., Snyder, C. R., and Davidson, B. J. (1980). Attention and the detection of signals. *J Exp Psychol*, 109(2):160–174.

Poustchi-Amin, M., Mirowitz, S. A., Brown, J. J., McKinstry, R. C., and Li, T. (2001). Principles and applications of echo-planar imaging: a review for the general radiologist. *Radiographics*, 21(3):767–779.

Robinson, D. L. and Petersen, S. E. (1992). The pulvinar and visual salience. *Trends Neurosci*, 15(4):127–132.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J.-D., Rees, G., Josephs, O., Deichmann, R., and Driver, J. (2006). Concurrent tms-fmri and psychophysics reveal frontal influences on human retinotopic visual cortex. *Curr Biol*, 16(15):1479–1488.

Schiller, P. H. and Lee, K. (1991). The role of the primate extrastriate area v4 in vision. *Science*, 251(4998):1251–1253.

Serences, J. T. and Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, 55(2):301–312.

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., and Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychol Sci*, 16(2):114–122.

Serences, J. T. and Yantis, S. (2007). Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. *Cereb Cortex*, 17(2):284–293.

Shipp, S. (2004). The brain circuitry of attention. *Trends Cogn Sci*, 8(5):223–230.

Shomstein, S. (2012). Cognitive functions of the posterior parietal cortex: top-down and bottom-up attentional control. *Front Integr Neurosci*, 6:38.

Siagian, C. and Itti, L. (2009). Biologically inspired mobile robot vision localization. *IEEE Transactions on Robotics*, 25(4):861–873.

Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J., and Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, 378(6556):492–496.

Snodderly, D. M. and Gur, M. (1995). Organization of striate cortex of alert, trained monkeys (macaca fascicularis): ongoing activity, stimulus selectivity, and widths of receptive field activating regions. *J Neurophysiol*, 74(5):2100–2125.

Stelzer, J., Chen, Y., and Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (mvpa): random permutations and cluster size control. *Neuroimage*, 65:69–82.

Thompson, K. G. and Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Prog Brain Res*, 147:251–262.

Thompson, K. G., Bichot, N. P., and Schall, J. D. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *J Neurophysiol*, 77(2):1046–1050.

Tootell, R. B., Hadjikhani, N. K., Vanduffel, W., Liu, A. K., Mendola, J. D., Sereno, M. I., and Dale, A. M. (1998). Functional analysis of primary visual cortex (v1) in humans. *Proc Natl Acad Sci U S A*, 95(3):811–817.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cogn Psychol*, 12(1):97–136.

Ts'o, D. Y., Frostig, R. D., Lieke, E. E., and Grinvald, A. (1990). Functional organization of primate visual cortex revealed by high resolution optical imaging. *Science*, 249(4967):417–420.

VanRullen, R. (2003). Visual saliency and spike timing in the ventral visual pathway. *J Physiol Paris*, 97(2-3):365–377.

Villringer, A. and Dirnagl, U. (1995). Coupling of brain activity and cerebral blood flow: basis of functional neuroimaging. *Cerebrovasc Brain Metab Rev*, 7(3):240–276.

Walther, D. and Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Netw*, 19(9):1395–1407.

Wandell, B. A., Dumoulin, S. O., and Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, 56(2):366–383.

Warnking, J., Dojat, M., Guérin-Dugué, A., Delon-Martin, C., Olympieff, S., Richard, N., Chéhikian, A., and Segebarth, C. (2002). fmri retinotopic mapping–step by step. *Neuroimage*, 17(4):1665–1683.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, 1(2):202–238.

Wolfe, J. M., Cave, K. R., and Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *J Exp Psychol Hum Percept Perform*, 15(3):419–433.

Zhang, X., Zhaoping, L., Zhou, T., and Fang, F. (2012). Neural activities in v1 create a bottom-up saliency map. *Neuron*, 73(1):183–192.

Zhaoping, L. and May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Comput Biol*, 3(4):e62.

Zipser, K., Lamme, V. A., and Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *J Neurosci*, 16(22):7376–7389.

6 RESEARCH ARTICLES

6.1 Orientation pop-out processing in human visual cortex

Bogler C., Bode S., Haynes J.D. (2013). Orientation pop-out processing in human visual cortex. *Neuroimage,* 81: 73-80.

NeuroImage 81 (2013) 73–80

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Orientation pop-out processing in human visual cortex

Carsten Bogler [a,b,c,*], Stefan Bode [b,c,d], John-Dylan Haynes [a,b,c,*]

[a] Bernstein Center for Computational Neuroscience Berlin, Charité-Universitätsmedizin Berlin, Germany
[b] Max-Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[c] Department of Neurology, Otto-von-Guericke University Magdeburg, Germany
[d] Melbourne School of Psychological Sciences, The University of Melbourne, Australia

ARTICLE INFO

Article history:
Accepted 6 May 2013
Available online 17 May 2013

Keywords:
Pop-out
Visual attention
Visual cortex
Functional MRI

ABSTRACT

Visual stimuli can "pop out" if they are different to their background. There has been considerable debate as to the role of primary visual cortex (V1) versus higher visual areas (esp. V4) in pop-out processing. Here we parametrically modulated the relative orientation of stimuli and their backgrounds to investigate the neural correlates of pop-out in visual cortex while subjects were performing a demanding fixation task in a scanner. Whole brain and region of interest analyses confirmed a representation of orientation contrast in extrastriate visual cortex (V4), but not in striate visual cortex (V1). Thus, although previous studies have shown that human V1 can be involved in orientation pop-out, our findings demonstrate that there are cases where V1 is "blind" and pop-out detection is restricted to higher visual areas. Pop-out processing is presumably a distributed process across multiple visual regions.

© 2013 Elsevier Inc. All rights reserved.

Introduction

A target that differs from distractors in its surround in a single elementary visual feature (such as luminance, color, orientation or motion) can easily be detected. For example a horizontal bar surrounded by many vertical bars perceptually "pops out". This pop-out effect is driven by the *feature contrast* between target and surround. Hence, in our example a vertical bar surrounded by horizontal bars would also pop out, and the pop-out would be stronger the higher the contrast between target and distractors is. The effect is maximal when the feature contrast is high or when the distractors are very similar, i.e. all distractors have exactly the same orientation or color. Under these conditions, visual search for pop-out targets is fast, does not require much attentional resources and is thought to run in parallel for different locations in the visual field, as opposed to serially scanning each location (Treisman and Gelade, 1980).

On a neural level, there has been some debate regarding the neural site of pop-out processing. One theory holds that the origin of the orientation pop-out effect is primary visual cortex (V1) (Kastner et al., 1997; Nothdurft et al., 1999; Zhang et al., 2012). Many studies have reported neural correlates for orientation selectivity in V1 in mammals and humans using a variety of methods including optical imaging (Ts'o et al., 1990), electrophysiological recordings (Hubel and Wiesel, 1962) as well as functional magnetic resonance imaging (fMRI) (Boynton and Finney, 2003; Tootell et al., 1998). Furthermore, the response of neurons in primary visual cortex to a particular stimulus can be modulated by stimuli in the non-classical receptive field. In other words, additional stimuli presented outside the classical receptive field of the respective neuron can influence the processing of the stimulus presented within the neuron's receptive field, for example by means of lateral inhibition (Knierim and van Essen, 1992; Li et al., 2000; Sillito et al., 1995; Zipser et al., 1996). These findings suggest that orientation pop-out could be processed in V1. Furthermore, there are some reports suggesting a direct representation of orientation pop-out in V1 (Kastner et al., 1997; Nothdurft et al., 1999; Zhang et al., 2012).

Other studies challenge this assumption. For example, Hegdé and Fellemann (2003) used a variety of different target–distractor configurations to investigate the response of V1 neurons to pop-out and non-pop-out stimuli. The authors demonstrated that neurons in V1 responded similarly to a target stimulus in their receptive field independent of whether it was embedded in a pop-out or a non-pop-out configuration. These results cast doubt on whether V1 could be the sole neural site for orientation pop-out. In line with this finding, a recent study (Burrows and Moore, 2009) demonstrated that neurons in V4 showed exactly the response profile that would be expected for a region that calculates orientation pop-out. V4 neurons showed increased firing rates only for targets that were surrounded by homogenous distractor sets, which were expected to create a pop-out effect, but not for other configurations of inhomogeneous distractor sets, which were expected to diminish the effect. Schiller and Lee (1991) investigated search performance on different displays after ablation of V4 in monkeys. On the side of the V4 lesion monkeys were severely limited in detecting a dark target between bright distractors and a small target between big distractors compared to the non-lesioned

* Corresponding authors at: Charité — Universitätsmedizin Berlin, Bernstein Center for Computational Neuroscience, Haus 6, Philippstrasse 13, 10115 Berlin, Germany.
E-mail addresses: carsten.bogler@bccn-berlin.de (C. Bogler), haynes@bccn-berlin.de (J.-D. Haynes).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.05.040

side. These results also suggest that V4 has an important role in processing pop-out stimuli and V1 alone is not sufficient to completely encode bottom-up saliency. Taken together, while some but not all studies show pop-out effects in V1, V4 might play a more crucial role for the pop-out calculation.

In the present study we aimed to further investigate the roles of V1, V4 and other potentially important brain areas in the calculation of orientation pop-out using fMRI. Unlike many other studies that treated pop-out as an all-or-nothing property, we experimentally manipulated pop-out parametrically by using different orientation contrasts between stimulus bars and surrounding distractor bars (0°, 30°, 60° and 90° differences) in a visual display. The pop-out effect was also measured behaviorally using reaction times in a separate psychophysical experiment. We then sought for brain regions with a neural response profile similar to the behavioral response profile in the fMRI experiment in which no overt responses were required and no confound with motor responses could occur. Finding such brain regions would strongly speak for a role in calculating the orientation contrast underlying the pop-out effect.

Methods

Psychophysical experiment

Participants

Twelve subjects (seven females, mean age 25.5 years, range 21–31) took part in the psychophysical study and gave written informed consent to the test procedure. The experiment was approved by the local ethics committee and was conducted according to the Declaration of Helsinki. All subjects were right-handed and had normal or corrected to normal visual acuity.

Visual stimuli and experimental procedure

The psychophysical experiment used a background of distractors consisting of a homogenous array of bars with a length of 2.2° (3 rows × 7 columns), all of which had the same orientation of either 0°, 45°, 90° or 135°. On each trial, one bar on either the left or right side, always displayed in the second row and the second column (top left) or sixth column (top right) was rotated counter-clockwise 15°, 30°, 45°, 60°, 75° or 90° relative to the distractor bars. The stimuli were presented on a 17-inch TFT-screen (resolution 1024 × 768, 60 Hz). The visual angle of the full display was $\alpha = 24.5° \times 19.7°$. The stimuli were presented for 500 ms with a fixed inter-stimulus-interval (ISI) of 2.5 s. Subjects had to fixate on a point displayed below the stimulus array in the lower center of the screen and indicated whether the position of the differently oriented bar was left or right by button press with the left and right index fingers respectively. Stimulus presentation and response recording were controlled using MATLAB 7.0 (The MathWorks, Inc.) in combination with the Cogent toolbox (http://www.vislab.ucl.ac.uk/Cogent). The distance between the fixation point and the center of the target bar was always 10.8° (Fig. 1).

Four experimental blocks (each 6 min duration) were conducted. During each block 144 trials (6 target orientations × 4 background orientations × 2 positions (left vs. right) × 3 repetitions) were presented. The aim of the behavioral experiment was to investigate whether the pop-out effect (expressed as faster reaction times for target detection) differed for the different target–distractor combinations. We expected that the reaction times decrease with increasing orientation contrasts.

Functional imaging experiment

Participants

Eleven subjects (6 females, mean age 28.7 years, range 24–34) took part in the neuroimaging study and gave written informed consent to the test procedure. Each subject participated in two scanning sessions on two different days. The experiment was approved by the local ethics committee and was conducted according to the Declaration of Helsinki. All subjects were right-handed and had normal or corrected to normal visual acuity.

Visual stimuli and experimental procedure

The visual display in the experiment consisted of a continuous stream of screens containing oriented bars. Between target displays the screens contained homogenous background stimuli (type A, see

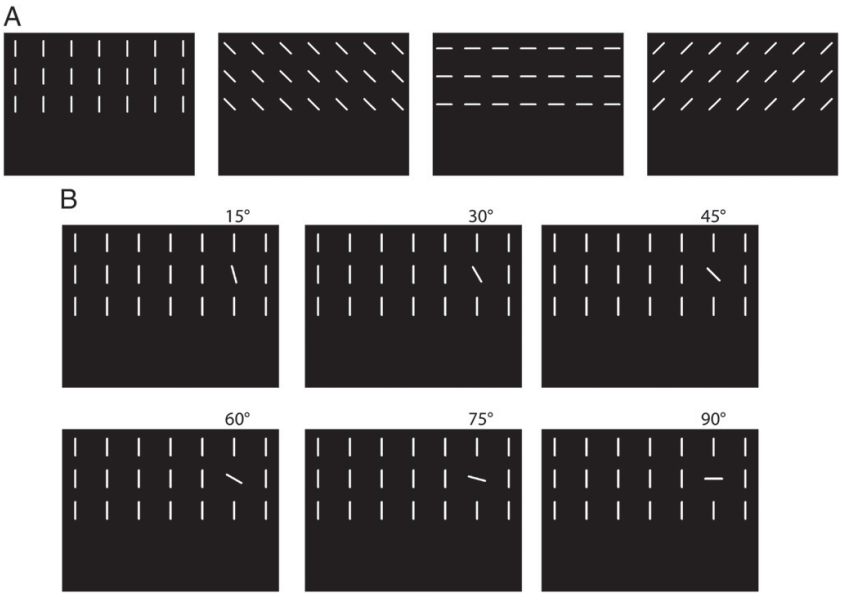

Fig. 1. Stimuli used in the behavioral experiment. A: Four different distractor configurations were used. B: A single bar on the left or right (shown in the example) side was rotated 15°, 30°, 45°, 60°, 75° or 90° counter clockwise to the distractor bars.

Fig. 2) showing an array of bars (3 rows × 7 columns) all of which had the same orientation of 0°, 45°, 90° or 135°. After every inter trial interval of 7, 10 or 13 s the screen switched to an orientation contrast target display that was the same as the background displays except for two target bars, one on the left and one on the right side (second row, second and sixth columns), which were rotated counter-clockwise 0°, 30°, 60° or 90° *relative* to the rest of the stimulus display (type B, see Fig. 2). The background displays were used as an implicit baseline in the GLM which only modeled the target display times (see below). Both target bars could be rotated in different angles on each trial. The combination of 4 different relative orientations on the left and right sides resulted in 16 conditions (Fig. 2 shows the matrix of possible combinations). Stimuli were presented via a LCD-projector (resolution 1024 × 768 pixel, 60 Hz) that projected from the head-end of the scanner onto a screen. Subjects viewed the projection through a mirror fixed on the head coil. The size of the display and viewing conditions were matched to the behavioral experiment. The visual angle of the full display was $\alpha = 26.6^{\circ} \times 20.5^{\circ}$. Orientation contrast stimulation of 4 s was alternated with 7, 10 or 13 s of homogenous background stimulation. Either homogenous background (type A) or orientation contrast (type B) stimuli were constantly flashed ON and OFF, with ON-phases corresponding to image presentation (500 ms) and OFF-phases corresponding to the presentation of a black background (500 ms). During each of the four-second orientation contrast stimulation periods, four different background orientations were alternated in a pseudo-randomized order while maintaining the relative orientation contrast the same at the position of the targets (see Fig. 2).

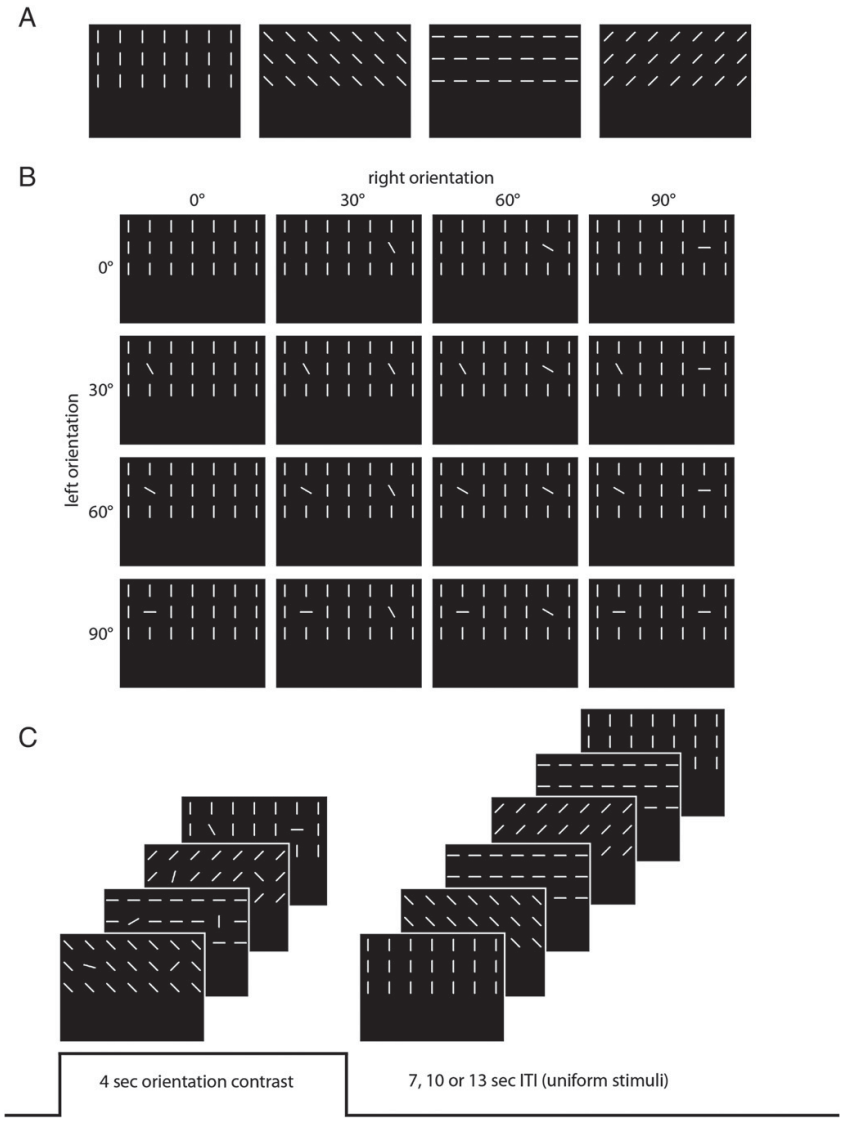

Fig. 2. Stimuli used in the fMRI-experiment. A: Homogenous stimuli. B: One bar on the left and right sides of the screen was rotated counter clockwise 0°, 30°, 60° or 90° relative to the distractor bars. B shows an example of all possible 16 conditions. C: A trial of the fMRI experiment. First 4 s of orientation contrast stimuli of one condition was presented. Importantly the 4 stimuli used during the stimulation have the same orientation contrast between the target and the distractor bars but different background orientations. After the orientation contrast stimulation homogenous stimuli were presented for 7, 10 or 13 s.

During the entire experiment a demanding fixation task was used to direct attention to the lower center of the screen where a small white outline square (visual angle $\alpha = 0.3° \times 0.3°$) was presented. Every 1000 ms, the square's left or right bar was removed for another 1000 ms. Subjects had to perform a one-back task and indicate by left or right button press whether the square "opened up" to the same side or the opposite side compared to the last opening. Four subjects were given a slightly different fixation task for the first session in which opening and closing of the square was twice as fast (500 ms) and the square could open to all four sides, not only to the left and the right. This version, however, appeared to be very demanding and was therefore modified for the following participants. Stimuli were presented and responses were recorded with MATLAB 7.0 (The MathWorks, Inc.) in combination with the Cogent toolbox (http://www.vislab.ucl.ac.uk/Cogent). During each scanning session, 5 runs of the experiment were conducted (each run lasted 672 s).

At the beginning of each scanning session a functional localizer for the positions of the two target bars was measured. Attention was always directed to the lower center of the screen using the same fixation task as in the main experiment. We presented a single bar (with identical size and position as the target bars of the main experiment) on either the right or left side of the screen. Bars were presented in blocks of 9.6 s, alternating between the left and right sides. Blocks were interrupted by no-stimulation periods for 4.8 s. During stimulation blocks the orientation of the bar alternated every 100 ms in a pseudo-randomized order (0°, 15°, 30°, 45°, 60°, 75°, 90°, 105°, 120°, 135°, 150° or 165°). The localizer scan lasted 345.6 s (12 repetitions of the left and right side stimulation). Four subjects were presented with a slightly different localizer in their first scanning session. A flickering checkerboard with a diameter corresponding to the target bars was shown at the same position. Blocks of flickering checkerboards (12 s) alternated left and right with no pause. These localizer scans lasted 288 s. Both versions of the functional localizer mapped the neural representation of the target bars equally well and we will refer to these areas irrespective of the localizer version.

fMRI acquisition

A Siemens TRIO 3 T scanner with standard head coil was used to acquire gradient-echo EPI functional MRI volumes covering the occipital, posterior temporal and parietal lobes (36 axial slices, TR = 2400 ms, echo time TE = 30 ms, resolution $3 \times 3 \times 2$ mm^3 with 1 mm gap). In each run, 280 images were acquired for each subject. For the functional localizer 2 runs with 144 images (120 images for localizer version 2) were acquired. The first two images were discarded to allow for magnetic saturation effects.

Table 1
Paired *t*-tests on the reaction times for all possible pair wise comparisons of the behavioral experiment.

Comparison	T-value	p-Value
15° vs. 30°	5.670	<0.001
15° vs. 45°	7.821	<0.001
15° vs. 60°	6.545	<0.001
15° vs. 75°	7.296	<0.001
15° vs. 90°	7.097	<0.001
30° vs. 45°	3.436	<0.01
30° vs. 60°	4.969	<0.001
30° vs. 75°	6.063	<0.001
30° vs. 90°	4.009	<0.01
45° vs. 60°	1.473	0.169 n.s.
45° vs. 75°	2.406	<0.05
45° vs. 90°	2.033	0.067 n.s.
60° vs. 75°	1.103	0.294 n.s.
60° vs. 90°	−.139	0.892 n.s.
75° vs. 90°	−.918	0.378 n.s.

Data analysis

Behavioral data

Hit rates and reaction times (correct trials only) were analyzed to investigate the behavioral effect of the 6 pop-out conditions.

Functional localizer and definition of regions of interest

Data from the independent functional localizer runs were smoothed with a 4 mm FWHM Gaussian kernel. We estimated a GLM with 2 regressors (visual stimulation left and visual stimulation right) that were convolved with a hemodynamic response function (HRF). The two normalized contrast maps for the two scanning sessions were averaged for each subject before entering a paired *t*-test.

Regions of interest (ROIs) were defined based on a combination of the functional localizer (e.g. left stimulation > right stimulation) and probabilistic cytoarchitectonic maps for early visual areas (Amunts et al., 2007).

Analysis of main experiment

We used three different analysis approaches in order to rule out the possibility that any negative findings were due to the specific methods used.

fMRI Analyses I and II: whole brain analysis

For the first two analyses the fMRI data were first motion corrected and then spatially smoothed with a Gaussian kernel of 6 mm FWHM using SPM2 (http://www.fil.ion.ucl.ac.uk/spm/). Data were highpass

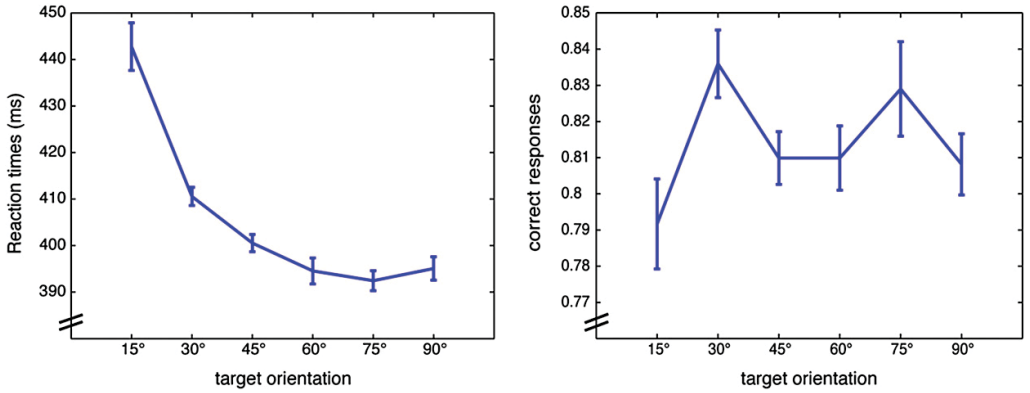

Fig. 3. Results of the behavioral experiment. Left: Reaction times for the 6 different pop-out conditions. Right: Hit rates for the different pop-out conditions.

Table 2
Paired *t*-tests on the hit rate for all possible pair wise comparisons of the behavioral experiment.

Comparison	T-value	p-Value
15° vs. 30°	−3.468	<0.01
15° vs. 45°	−1.450	0.175 n.s.
15° vs. 60°	−1.113	0.289 n.s.
15° vs. 75°	−1.780	0.103 n.s.
15° vs. 90°	−1.059	0.312 n.s.
30° vs. 45°	2.472	<0.05
30° vs. 60°	2.123	0.057 n.s.
30° vs. 75°	0.454	0.659 n.s.
30° vs. 90°	1.810	0.098 n.s.
45° vs. 60°	0.000	1.000 n.s.
45° vs. 75°	−1.554	0.148 n.s.
45° vs. 90°	0.124	0.903 n.s.
60° vs. 75°	−1.246	0.239 n.s.
60° vs. 90°	0.161	0.875 n.s.
75° vs. 90°	1.127	0.284 n.s.

Table 4
Main effect of pop-out stimulation. The coordinates are given according to MNI space with their T-values. L = left hemisphere, R = right hemisphere; all $p < 0.0001$ uncorrected, voxel extend threshold 10.

Anatomical area	L/R	T-value	X	Y	Z
V1	R	4.77	3	−75	6
	L	4.39	−6	−72	3
V3v; V4	R	13.53	24	−66	−6
	L	12.02	−21	−69	−12
MT+; V5	R	7.17	54	−75	9
	L	8.57	−27	−81	21
Fusiform gyrus; inferior temporal gyrus	R	4.83	45	−51	−12
	L	5.80	−51	−54	−15
Intraparietal sulcus; hIP3; hIP1	R	4.18	30	−63	51
	L	4.67	−24	−60	39
Right pallidum	R	4.72	21	−6	0
Right inferior frontal gyrus	R	4.90	45	33	−3
Right thalamus	R	5.12	12	−21	9
N/A; left thalamus	L	4.73	−9	−3	−9

filtered with a cut-off period of 128 s. We then modeled the cortical response to the pop-out stimuli using two different GLMs.

(I) A GLM with 16 HRF-convolved regressors was estimated using the four orientation differences between target and distractors on the left and the right sides ($4 \times 4 = 16$ conditions).
(II) A GLM with one HRF-convolved regressor and 2 additional parametric regressors were estimated (Friston et al., 1997). The two parametric regressors reflected the orientation difference between target and distractors (1 to 4 corresponding to 0° to 90°) for the left or right side respectively.

In both analyses the stimulation with the homogenous stimuli served as an implicit baseline condition. The resultant contrast maps were normalized to a standard stereotaxic space (Montreal Neurological Institute EPI template) and re-sampled to an isotropic spatial resolution of $3 \times 3 \times 3$ mm^3. For each subject normalized contrast maps were averaged across the two scanning sessions before random effects general linear models (ANOVA) were estimated across subjects. The first ANOVA was based on the first GLM and searched for a linear trend for increasing orientation pop-out using the contrast [−1.5 −0.5 0.5 1.5]. The second ANOVA was based on the parametric model and specifically searched for voxels in which the BOLD response was significantly modulated by relative difference between target and background bars. We additionally implemented a non-linear model for the whole brain BOLD analysis:

$$y = 1 - \frac{1}{x}$$

(with x = {1 2 3 4} representing 0°, 30°, 60° and 90° orientation contrasts and y the expected BOLD response). This model provided stronger correlation with the reaction times for 30°, 60° and 90° orientation contrasts (r = −0.92, as compared to r = −0.85 for the initial linear model), thus it produced a closer match to the behavioral data.

Table 3
Main effect of the functional localizer. The coordinates are given according to MNI space with their T-values. L = left hemisphere, R = right hemisphere; all $p < 0.001$ uncorrected, voxel extend threshold 10.

Anatomical area	L/R	T-value	X	Y	Z
Stimulation left vs. right					
V1; V2	R	7.97	6	−72	6
V4; V3v	R	7.62	24	−63	−9
MT+	R	8.69	48	−69	6
Stimulation right vs. left					
V1; V2	L	7.27	−6	−75	3
V4; V3v	L	9.80	−18	−69	−12
MT+	L	8.05	−51	−78	9
N/A	L	7.52	−24	−81	18

* Bonferroni corrected significance level $p < 0.01$.

fMRI Analysis III: ROI analysis

For the third analysis we estimated the GLM with 16 conditions (GLM I) but now based on realigned but unsmoothed data. Parameter estimates were extracted from locations defined by the localizer-based ROIs (see above) and averaged for each ROI. For each subject the correlation between the parameter estimates in the ROIs and the four levels of orientation pop-out [0, 0.5, 0.67, 0.75; assuming a non-linear increase as described above] were calculated for each session, Fisher-Z normalized and then averaged across the two sessions. Finally, a random effects analysis was conducted to statistically test the Fisher-Z normalized correlation coefficients for each ROI.

Results

The reaction times (RTs) and accuracy for the different pop-out conditions can be seen in Fig. 3. RTs for pop-out on the left and right visual fields were averaged within conditions because there was no difference between sides: A repeated measures ANOVA with the factors side (left, right) and pop-out (15°, 30°, 45°, 60°, 75° and 90° orientation differences between target and distractors) showed a main effect for pop-out ($F(2.03, 22.34) = 35.41$, $p < 0.0001$) but not for side. The main effect for pop-out remained unchanged after averaging ($F(2.02, 22.21) = 35.29$, $p < 0.0001$). Separate *t*-tests confirmed the increase in RTs with decreasing orientation contrast (all tests can be found in Table 1). Furthermore, a linear regression analysis confirmed the significant negative linear trend in the RT data ($b = -.57$, $t(70) = -3.15$, $p < 0.01$, $R^2 = .12$, $F(1, 70) = 9.93$, $p < 0.01$). Hit rates for correct identification did not differ between conditions ($F(3.26, 35.82) = 2.38$, $p = 0.08$; separate *t*-tests can be found in Table 2) and no linear trend was found ($b = 0$, $t(70) = 0.69$, $p = .49$, $R^2 = .01$, $F(1, 70) = 0.48$, $p = .49$).

All subjects confirmed in debriefing interviews that the fixation task was demanding and required continuous allocation of attention. The mean hit rate for the fixation task was 89.6%. One subject misunderstood the fixation task, left out every second response and thus was not included in the calculation of the mean performance.

Analysis I (implementing 16 conditions: 4×4 orientation contrast combinations on the left and right sides) revealed an increased BOLD response with the onset of the orientation contrast stimulation (collapsed across all relative orientations) in V1, V4, MT+ and IPS ($p < 0.05$, FDR corrected, see Table 4 for all regions). V1, V4 and MT+ overlapped with the ROIs from the independent functional localizer (Figs. 4A and B). The only region, however, showing a graded and systematic modulation of the BOLD response by level of pop-out in the

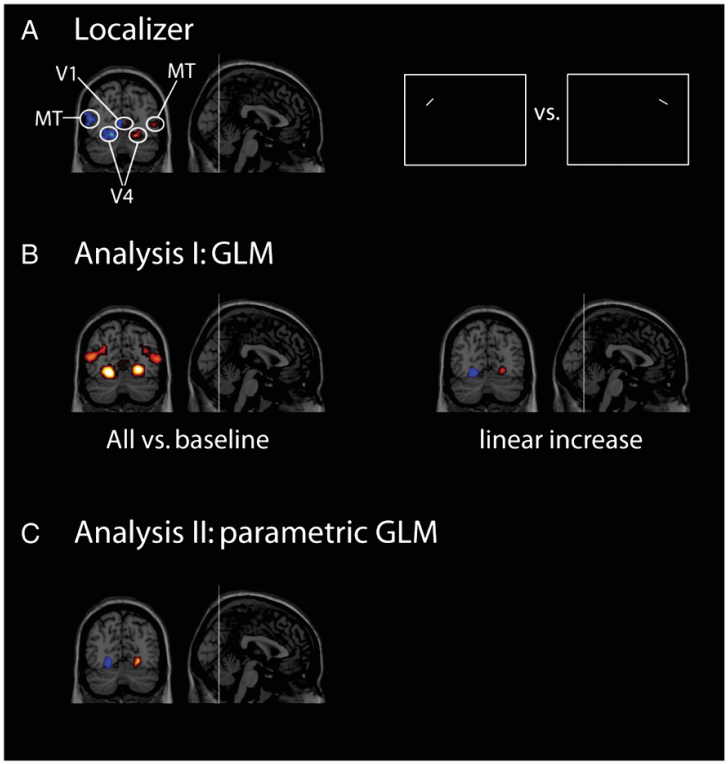

Fig. 4. Results of Analyses I and II. A: Functional localizer; visual stimulation left vs. visual stimulation right (red) and visual stimulation right vs. visual stimulation left (blue) ($p < 0.001$ uncorrected, cluster threshold = 10 voxel). B left: Main effect for pop-out stimulation vs. baseline ($p < 0.0001$ uncorrected, cluster threshold = 10 voxel). See Tables 3 and 4 for a complete list of all activated regions. B right: Linear increase for level of pop-out left and right ($p < 0.0001$ uncorrected, cluster threshold = 10 voxel). C: Results of Analysis II ($p < 0.0001$ uncorrected, cluster threshold = 10 voxel).

contralateral hemifield was bilateral V4 ($p < 0.05$, FDR corrected). This region showed an increase of activation with increasing orientation difference between target and distractors.

The parametric GLM analysis (Analysis II), which modeled the linear increase of BOLD with increasing pop-out directly, confirmed the effect for contralateral stimulation in V4 ($p < 0.0001$ uncorrected, cluster size > 10 voxels) (Fig. 4C). No other region could be found to show a parametric effect of pop-out.

Analyses I and II were also conducted using a non-linear response model. Both analyses again showed a graded and systematic modulation of the BOLD response by level of pop-out (now expecting a non-linear relationship between BOLD and orientation contrast) in contralateral V4 (Fig. 5A). The analyses with the non-linear response model fitted the data slightly better than the respective analyses based on the linear model (ANOVA with linear T-contrast: 5.01 (left V4)/6.01 (right V4); ANOVA with non-linear T-contrast: 5.77 (left V4)/6.01 (right V4); parametric GLM with linear response model: 7.49 (left V4)/5.25 (right V4); parametric GLM with non-linear response model: 8.05 (left V4)/5.63 (right V4)).

For Analysis III we first defined ROIs based on an independent functional localizer combined with anatomical ROIs. The functional localizers revealed increased BOLD response in striate (V1) and extrastriate (V4 and MT+) visual cortex contralateral to the stimulation side ($p < 0.001$, cluster size > 10 voxels) (Fig. 4A). Activity of the functional localizer combined with anatomical ROIs served as ROIs for the analysis of the main experiment. The ROI analyses confirmed that the V1 ROI did not show any systematic modulation for the different pop-out conditions (see Fig. 5 and Table 5). The V3 and V4 ROIs, on the other hand, showed a significant effect for orientation pop-out (see Table 5).

Stimulation with different orientation contrasts involved pop-out stimuli (30°, 60° and 90° orientation contrasts) and non pop-out stimuli (0° orientation contrast). To investigate whether any regions showed increased activity for pop-out compared to non-pop-out stimuli we averaged the responses to the respective conditions and compared pop-out against non-pop-out stimulation using a whole brain analysis as well as a ROI analysis for the visual areas. The whole brain analysis of pop-out compared to non-pop-out stimulation showed increased activity only in V4 contralateral to the stimulated hemifield ($p < 0.05$ FWE corrected, Fig. 5C). The difference between averaged parameter estimates for pop-out vs. non-pop-out within the five ROIs marginally missed the significance threshold after correcting for multiple comparisons in V3 and V4 (V3: $T(10) = 2.62$ $p = 0.03$; V4: $T(10) = 2.98$ $p = 0.01$; Bonferroni corrected significance level $p < 0.01$, Fig. 5D).

Discussion

In the present study using fMRI we found a representation of orientation contrast in extrastriate visual cortex (V4) that is likely to underlie the visual pop-out effect. V4 was the only visual region with a reliable pop-out representation. Importantly, V1 did not encode information about the orientation contrast, as confirmed using a combination of different models and approaches.

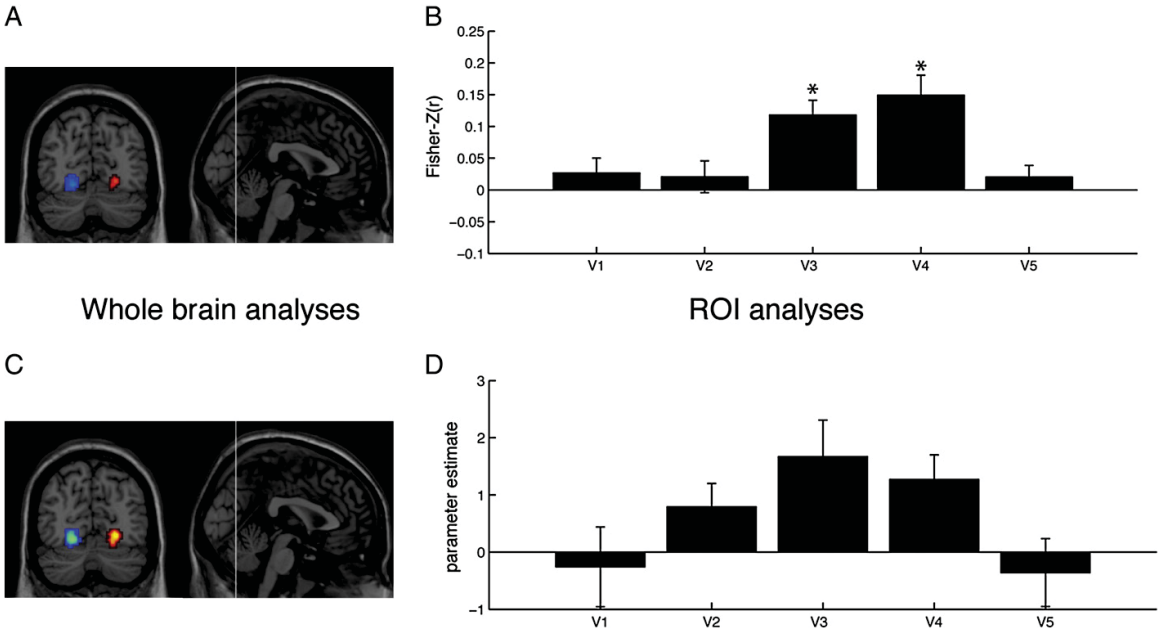

Fig. 5. A: Whole brain analysis for non-linear increase for level of pop-out left and right ($p < 0.0001$ uncorrected, cluster threshold = 10 voxel). B: Fisher-Z normalized and averaged correlation coefficients between the averaged parameter estimates within the five different ROIs and the relative orientation difference. C: Whole brain analysis of pop-out (30°, 60° and 90° orientation contrast) vs. no pop-out (0° orientation contrast) showed increased activity only in V4 contralateral to the stimulated hemifield ($p < 0.05$ FWE corrected; $p < 0.001$ uncorrected, voxel extend threshold 10 voxels for illustration purpose). D: Difference between averaged parameter estimates for pop-out vs. non-pop-out within the five different ROIs. Only V3 and V4 are close to corrected significance level (V3: $T(10) = 2.62$ $p = 0.03$; V4: $T(10) = 2.98$ $p = 0.01$; Bonferroni corrected significance level $p < 0.01$).

We found that V3 and V4 were the only regions in which the neural response to pop-out stimuli clearly mirrored behavioral responses to the same stimuli in the complementary reaction time experiment. One possible conclusion from these findings is that V4 directly encodes pop-out as in a previous study (Bogler et al., 2011) in which we identified representations of graded saliency in striate and extrastriate visual cortex. In this former study we used natural scenes in combination with a computational model that combined multiple features, not just orientation contrast. There is overlap between the activated cortical regions for saliency processing in the previous and orientation pop-out in the current study, but in the previous study both striate and extrastriate visual cortex were activated. This could be explained by the rich natural stimuli that were presented which contained more features that could be processed for the calculation of a saliency map. In contrast, our current study demonstrates that under specific stimulus conditions pop-out processing and V1 responses can be dissociated.

In the present study we further showed that there was no representation of orientation contrast in V1. This does not rule out the existence of orientation contrast information in V1 as our methods could lack the required sensitivity for V1; however, please note that the pop-out signals we find in V4 are highly robust; furthermore we do find clearly significant and localized responses in V1 for the orientation contrast stimulation in general but these responses were not modulated by the level of pop-out. Please note, that there was also no difference in activity between pop-out (i.e. 30°, 60°, and 90° orientation contrast) and non-pop-out stimuli (0°) in V1 (Figs. 5C, D).

Table 5
Results of the ROI analysis. *t*-Test on Fisher-Z normalized correlation coefficients between BOLD-response averaged within the ROI and orientation-contrast.

		p	T(10)
ROI	V1	0.2797	1.1429
	V2	0.4246	0.8325
	V3	0.0004*	5.1455
	V4	0.0007*	4.7552
	V5/MT+	0.2884	1.1212

* Bonferroni corrected significance level $p < 0.01$.

In contrast, a recent study from Zhang et al. (2012) found a representation of orientation pop-out in V1, V2, V3 and V4. Furthermore, they found that the correlation between a psychophysical measure for pop-out and the BOLD response was largest in V1. In other words their pattern of results is inverse to ours. One explanation for this could be the difference between the studies in stimulus size. Zhang and colleagues used 0.75° bars while our bars were 2.2° in size. Hegdé and Felleman (2003) reported mean receptive field sizes of V1 neurons of 1.2° similar to what have been previously reported (Essen and Zeki, 1978; Snodderly and Gur, 1995). The receptive field size of the V4 neurons in the study of Burrows and Moore (2009) was between 2.5° and 5°. So it might be that our stimuli were rather optimized for V4 and not for V1.

The involvement of early versus higher visual areas in attentional processing is known to depend on spatial scale (Hopf et al., 2006; Kastner et al., 2001). Importantly, we still obtained a clear psychophysical effect with the used stimuli. However, this means that some stimuli cannot be represented in a saliency map in V1. Therefore, V1 cannot act as a sole stimulus-independent saliency map as suggested by Zhang et al. (2012). A saliency map could instead be hierarchically organized, as suggested by VanRullen (2003). This would also explain how very salient but complex objects such as faces (Cerf et al., 2009) are encoded in a saliency map that is distributed through the ventral visual pathway.

Another explanation for our results could be found in normalization models. In general, a potential mechanism of orientation pop-out could be the orientation-tuned surround suppression that is strongest for

collinear surround stimuli and weakest for orthogonal surround stimuli (Cavanaugh et al., 2002; Knierim and van Essen, 1992). This could be explained by a normalization model (Carandini and Heeger, 2012) with a normalization pool that exerts an orientation-tuned surround inhibition. Our finding of a pop-out signal in V4 but not in V1 could reflect a dependence of normalization on spatial scale. The larger receptive fields in V4 might integrate information from a normalization pool (Carandini and Heeger, 2012) covering larger regions of the visual field.

Taken together, the results of our study show that orientation contrast for salient stimuli is represented in V4. We conclude that saliency processing is a rather distributed process, involving large parts of the visual cortex as opposed to being merely localized in V1.

Acknowledgments

This work was funded by the Bernstein Computational Neuroscience Program of the German Federal Ministry of Education and Research (BMBF grant 01GQ0411), the Bernstein Focus Neurotechnology (BMBF grant 01GQ0851), the German Research Foundation (DFG HA 5336/1-1), the Excellence Initiative of the German Federal Ministry of Education and Research (DFG grant GSC86/1-2009), and the Max Planck Society.

Conflict of interest

There is no conflict of interest.

References

Amunts, K., Schleicher, A., Zilles, K., 2007. Cytoarchitecture of the cerebral cortex — more than localization. NeuroImage 37 (4), 1061–1065 (discussion 1066–8).
Bogler, C., Bode, S., Haynes, J.-D., 2011. Decoding successive computational stages of saliency processing. Curr. Biol. 21 (19), 1667–1671.
Boynton, G.M., Finney, E.M., 2003. Orientation-specific adaptation in human visual cortex. J. Neurosci. 23 (25), 8781–8787.
Burrows, B.E., Moore, T., 2009. Influence and limitations of popout in the selection of salient visual stimuli by area v4 neurons. J. Neurosci. 29 (48), 15169–15177.
Carandini, M., Heeger, D.J., 2012. Normalization as a canonical neural computation. Nat. Rev. Neurosci. 13 (1), 51–62.
Cavanaugh, J.R., Bair, W., Movshon, J.A., 2002. Selectivity and spatial distribution of signals from the receptive field surround in macaque v1 neurons. J. Neurophysiol. 88 (5), 2547–2556.
Cerf, M., Frady, E.P., Koch, C., 2009. Faces and text attract gaze independent of the task: experimental data and computer model. J. Vis. 9 (12), 10.1–10.1015.
Essen, D.C., Zeki, S.M., 1978. The topographic organization of rhesus monkey prestriate cortex. J. Physiol. 277, 193–226.
Friston, K., Price, C., Buchel, C., Frackowiak, R., 1997. A taxonomy of study design. In: Frackowiak, R., Friston, K., Frith, C., Dolan, R., Mazziotta, J. (Eds.), Human brain function. Academic Press USA, pp. 141–159.
Hegdé, J., Felleman, D.J., 2003. How selective are v1 cells for pop-out stimuli? J. Neurosci. 23 (31), 9968–9980.
Hopf, J.-M., Luck, S.J., Boelmans, K., Schoenfeld, M.A., Boehler, C.N., Rieger, J., Heinze, H.-J., 2006. The neural site of attention matches the spatial scale of perception. J. Neurosci. 26 (13), 3532–3540.
Hubel, D.H., Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160, 106–154.
Kastner, S., Nothdurft, H.C., Pigarev, I.N., 1997. Neuronal correlates of pop-out in cat striate cortex. Vis. Res. 37 (4), 371–376.
Kastner, S., Weerd, P.D., Pinsk, M.A., Elizondo, M.I., Desimone, R., Ungerleider, L.G., 2001. Modulation of sensory suppression: implications for receptive field sizes in the human visual cortex. J. Neurophysiol. 86 (3), 1398–1411.
Knierim, J.J., van Essen, D.C., 1992. Neuronal responses to static texture patterns in area v1 of the alert macaque monkey. J. Neurophysiol. 67 (4), 961–980.
Li, W., Thier, P., Wehrhahn, C., 2000. Contextual influence on orientation discrimination of humans and responses of neurons in v1 of alert monkeys. J. Neurophysiol. 83 (2), 941–954.
Nothdurft, H.C., Gallant, J.L., Essen, D.C.V., 1999. Response modulation by texture surround in primate area v1: correlates of "popout" under anesthesia. Vis. Neurosci. 16 (1), 15–34.
Schiller, P.H., Lee, K., 1991. The role of the primate extrastriate area v4 in vision. Science 251 (4998), 1251–1253.
Sillito, A.M., Grieve, K.L., Jones, H.E., Cudeiro, J., Davis, J., 1995. Visual cortical mechanisms detecting focal orientation discontinuities. Nature 378 (6556), 492–496.
Snodderly, D.M., Gur, M., 1995. Organization of striate cortex of alert, trained monkeys (*Macaca fascicularis*): ongoing activity, stimulus selectivity, and widths of receptive field activating regions. J. Neurophysiol. 74 (5), 2100–2125.
Tootell, R.B., Hadjikhani, N.K., Vanduffel, W., Liu, A.K., Mendola, J.D., Sereno, M.I., Dale, A.M., 1998. Functional analysis of primary visual cortex (v1) in humans. Proc. Natl. Acad. Sci. U. S. A. 95 (3), 811–817.
Treisman, A.M., Gelade, G., 1980. A feature-integration theory of attention. Cogn. Psychol. 12 (1), 97–136.
Ts'o, D.Y., Frostig, R.D., Lieke, E.E., Grinvald, A., 1990. Functional organization of primate visual cortex revealed by high resolution optical imaging. Science 249 (4967), 417–420.
VanRullen, R., 2003. Visual saliency and spike timing in the ventral visual pathway. J. Physiol. Paris 97 (2–3), 365–377.
Zhang, X., Zhaoping, L., Zhou, T., Fang, F., 2012. Neural activities in v1 create a bottom-up saliency map. Neuron 73 (1), 183–192.
Zipser, K., Lamme, V.A., Schiller, P.H., 1996. Contextual modulation in primary visual cortex. J. Neurosci. 16 (22), 7376–7389.

6.2 Dissociation between saliency signals and activity in early visual cortex

Betz T., Wilming N., Bogler C., Haynes J.D., König P. (submitted to Journal of Vision). Dissociation between saliency signals and activity in early visual cortex.

Dissociation between saliency signals and activity in early visual cortex

Torsten Betz[1,2,3+], Niklas Wilming[1+], Carsten Bogler[2,4], John-Dylan Haynes[2,4*], Peter König[1,5*]

[1]Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
[2]Bernstein Center for Computational Neuroscience Berlin, Charité – Universitätsmedizin Berlin, Germany
[3]Berlin Institute of Technology, Berlin, Germany
[4]Berlin Center for Advanced Neuroimaging, Charité – Universitätsmedizin Berlin
[5]Dept. of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[*] shared senior authorship
[+] shared first authorship

Abstract

Saliency is a measure that describes how attention is guided by local stimulus properties. Some hypotheses assign its computation to specific topographically organized areas of early human visual cortex. However, in most stimuli, saliency is correlated with luminance contrast, which in turn is known to correlate with activity in these early areas. Thus, any observed correlation of local activity with saliency might be due to the area encoding luminance contrast. Here we disentangle encoding of local luminance contrast and saliency by using stimuli where the two properties are uncorrelated. First we conduct an eye-tracking study to verify that both negative and positive contrast modifications located in individual quadrants of the visual field increase saliency. Second, subjects view identical stimuli while fMRI BOLD signals are recorded. We find that positive contrast modifications induce a robust increase of activity in V1-V4. However, negative contrast modifications lead to a reduced activity level compared to unmodified quadrants. Furthermore, even with multivariate pattern classification techniques it is not possible to decode the location of the salient quadrant independent of the type of the contrast modification. Instead, decoding of the contrast-modified location is only possible separately for the two modification types. These findings suggest that the BOLD activity in early visual areas is dominated by contrast-dependent processes, and does not include the contrast-invariance necessary for the computation of feature-invariant saliency.

Keywords: visual attention, saliency, visual cortex, functional MRI, eye-tracking

Introduction

The brain continuously samples information by directing covert or overt attention towards different locations in the environment. In recent years the underlying mechanisms of attentional selection have moved towards the center of research interest. An early hypothesis suggests that the brain computes a saliency map – a topographically organized representation of the visual field that can be used to decide where to attend next. The more salient a position is, the more likely it will be attended (Koch & Ullman, 1985).

This hypothesis has triggered the search for brain areas that perform the required computations and encodes saliency. Several regions in cerebral cortex and subcortical areas have been suggested as the locus of a saliency map: superior colliculus (Kustov & Robinson, 1996), pulvinar (Shipp, 2004), V1 (Li, 1999; Li, 2002), parietal cortex (Bisley & Goldberg, 2010; Geng & Mangun, 2009; Gottlieb, Kusunoki, & Goldberg, 1998; Serences et al., 2005), V4 (Mazer & Gallant, 2003) and frontal eye fields (Serences & Yantis, 2007; Thompson & Bichot, 2005). The term saliency is often used for task-dependent as well as stimulus-driven processes that might occur in different areas, an ambiguity that may lead to confusion. Here, we are only concerned with the latter. Li Zhaoping (Li, 2002; Zhang et al., 2012; Zhaoping, 2011) argues that already V1 creates a purely stimulus driven (bottom up) saliency map that relays information to higher areas.

In a network of interacting areas it is difficult to identify and locate computations of saliency. Because V1 projects directly or indirectly to all areas listed above, it is reasonable to assume that saliency information would also be observable in these other areas if it is already computed in V1 (Shipp, 2004; Zhang et al., 2012). Any higher cortical area that receives information from many other areas is therefore likely to exhibit saliency map like properties because saliency related information propagates up the hierarchy. To identify where saliency information is first made explicit, as opposed to simply received, it is therefore important to identify the exact contribution of early visual areas, like V1, to this computation.

A complication arises from the fact that early areas are usually considered to encode certain features of the stimulus (e.g. oriented edges in V1, Hubel & Wiesel, 1959), and these features in turn contribute to the saliency map (Itti & Koch, 2001). Thus, a strong response to a salient stimulus in an area need not reflect the explicit computation of saliency but potentially only the encoding of a feature that also influences the computation of saliency. For example, high contrast edges strongly activate V1 neurons and high contrast edges correlate with saliency. But encoding of high contrast edges in V1 does not necessarily form an explicit representation of saliency.

In this study, we investigate the contribution of early visual areas (V1-V4) to the computation of visual saliency. We address the dependency of contrast and saliency by exploiting a finding by Einhäuser & König (2003): Under certain conditions a local reduction of luminance contrast leads to an increase in saliency. A brain region that explicitly encodes saliency would show an increased activity in response to local contrast reductions (saliency encoding hypothesis). However, a brain region that encodes contrast would show reduced activity (contrast encoding hypothesis). Hence, we created stimuli in which the luminance contrast in one of the four quadrants was either increased or decreased. An eye tracking study confirms that both contrast modifications increased saliency. We then present these stimuli in an fMRI experiment and record BOLD responses. The type of representation is characterized by analyzing the mean BOLD activity and by multivariate pattern classification in functionally defined regions of interest (ROI V1-V4). This allows differentiating between encoding of luminance and encoding of saliency.

Methods

Stimuli

We use a set of pink noise images as stimuli to avoid the influence of high level factors and still retain some of the statistics of natural stimuli. We generated these stimuli by randomizing the phases in Fourier transformed natural images, which removes all image structure, but leaves the power spectrum untouched (Einhäuser et al., 2006). 27 source images were chosen randomly from the categories "natural" and "manmade" (Açık, Sarwary, Schultze-Kraft, Onat, & König, 2010). Their luminance histograms were flattened, and contrast increases and decreases were applied to each quadrant. Contrast modifications were computed as described before (Açik, Onat, Schumann, Einhäuser, & König, 2009). Modified luminance values were given by:

$$I = [1+\alpha K]*[I_0 - M] + M$$

where I_0 denotes the source image, M the mean image and α the peak modification level. * is a pointwise multiplication of matrices. α was set to 0.9 for contrast increases and to -0.9 for contrast decreases. I, K and M are matrices of the same size as I_0. M contains the local mean luminance values and is computed by convolving I_0 with a Gaussian kernel with a full width at half maximum of 6.53°. K describes how the modification level changes as a function of the distance to the peak modification. Here, K is given as the cosine of the square of the distance to the peak modification location.

$$K(x,y)=(\cos[(x^2+y^2)/s]+1) / 2$$

This function has its maximum value of 1 at x=y=0, and smoothly drops to 0 at $(x^2+y^2)/s = \pi$. s was chosen such that the zero-crossing occurred at the edge of the modified quadrant in the horizontal direction. In the vertical direction, the modification slightly leaks into the adjacent quadrant, but this only corresponds to 0.74% of the kernel's mass. To modify a specific quadrant, K was centered in the respective quadrant. All in all we generated 243 different stimuli (27 stimuli * 2 modifications * 4 quadrants + 27 unmodified; see Figure 1 for examples).

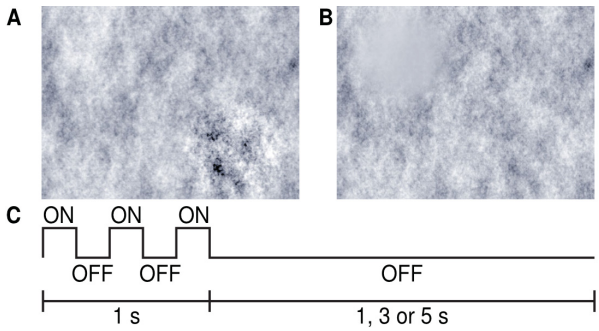

Figure 1. A) and B) Examples of pink noise stimuli with high contrast and low contrast modifications. Note that the change in contrast is much more gradual and less visible if stimuli are viewed at their original size. C) Time-course of an fMRI trial. In each trial, one image was presented repeatedly for 200 ms with a 200 ms gap. Between successive trials there was a variable interstimulus interval of 1 to 5 s.

Eyetracking

Participants

Eleven student volunteers took part in the eye tracking experiment (4 male and 7 female, age range 20-30 years, mean age 25 years). All participants had normal or corrected to normal vision. Inclusion in the study was contingent on reliable eye tracking calibration with an average validation error below 0.3°. As compensation, participants received either payment or course credit. The study was conducted in accordance with the declaration of Helsinki, and approved by the local ethics committee.

Apparatus

The experimental apparatus was designed to resemble the parameters of the fMRI experiment. Screen distance was 80 cm to achieve the same stimulus size as later on in the scanner (26.6° x 20.5°). Stimuli were presented on a 19-inch flat screen monitor (SyncMaster 971p, Samsung Electronics, Seoul, South Korea) at a screen resolution of 1280x1024 pixel and refresh rate of 75Hz. Eye movements were recorded with an Eyelink II system (SR Research Ltd., Mississuaga, Ontario, Canada) at 500Hz sampling rate. The system is capable of tracking both eyes, but only the eye that gave a lower validation error after calibration was recorded. A chin-rest was used to minimize head movements. The experiment was conducted in a darkened room.

Procedure

The stimuli were presented in 3 blocks of equal length. In the break between blocks we encouraged participants to rest and remove the eye tracker. Before each stimulus onset drift correction was performed, requiring participants to fixate the center of the screen. Subsequently, each stimulus was presented until a random number of saccades between three and eight had been performed. The stimulus order was randomized, but the number of stimuli from each condition was the same in all blocks. The task was to recognize whether a patch of size 250x250 pixels, presented after stimulus offset, was taken from the image just shown. The probability that the patch actually came from the previously seen image was 50%. Participants responded by pressing either the “arrow up” button for “yes” or the “arrow down” button for “no” on a regular keyboard. To shorten the duration of the experiment and to avoid fatigue or demotivation, the patch recognition task was only presented after 49 randomly selected trials. A test run, consisting of five images, was performed in order to let the subjects gain experience with the task.

fMRI

Participants

14 naive observers participated in the fMRI study (8 female, 6 male; age range 22-33 years, mean age 27 years). They reported normal or corrected to normal vision, and received payment for their participation. The study was conducted in accordance with the declaration of Helsinki, and approved by the local ethics committee. For all participants, detailed anatomical brain images were available from previous studies. The data of two observers had to be excluded from analysis. One fell asleep during the experiment, and for the other, retinotopic mapping was not successful.

Apparatus

Images were presented with a Sanyo Xtra Pro (SANYO Electric Co., Ltd., Osaka, Japan) with a resolution of 1024x768 pixel. Only the central 800x600 pixel (26°x20°) were in clear view, so images were resized to this resolution with bicubic interpolation. A Siemens 3T Magnetom (Siemens AG, Erlangen, Germany) was used to acquire functional MR EPI volumes with 36 slices at an isotropic resolution of 3 mm^3 (TR = 2000ms; TE=24ms; 36 axial slices; FOV 192x192x108 mm; α = 0°). Structural images were acquired with a T1-weighted 3D MP-RAGE with selective water excitation and linear phase encoding. Magnetization preparation consisted of a nonselective inversion pulse. The imaging parameters were TI = 650 ms, TR = 1300 ms, TE = 3.93 ms, α = 10°, spatial resolution 1 mm^3 isotropic, two averages.

Procedure

The pink noise experiment was divided into five different runs of stimulus presentations. Between each run, participants could take a small break and relax their eyes. Each run lasted approximately 10 minutes and consisted of 108 presentations of pink noise stimuli. Each run contained an equal number of stimuli from each condition, interleaved with 30 blank trials. In contrast to the eye tracking experiment, a task was given that required fixating the center of the screen. The outline of a 0.3°x0.3° square was drawn in black in the middle of the screen. Every 1200ms either the left or right side of the square opened for 600ms and participants had to indicate which side was opened by pressing one of two buttons with the middle or index finger of their right hand. In each trial, one pink noise stimulus was flashed three times for 200ms in the background, with a 200ms gap between flashes. The inter-trial interval was variable, 1s, 3s or 5s (see Figure 1C). The fixation task was independent of the stimulus presentation. 286 functional MRI volumes were acquired in each run. A 42-slice whole brain EPI image was also acquired to facilitate spatial normalization.

Retinotopic mapping and localization runs were conducted in a separate session on a different day. The retinotopic mapping runs consisted of the presentation of a rotating wedge (5cyc/300s) and an expanding ring (10expansions/300s). This allowed for a functional definition of early visual areas, especially V1-V4 (Wandell, Dumoulin, & Brewer, 2007; Warnking, 2002). To localize voxels that react

to visual stimulation of a quadrant, flickering checkerboard patterns that were centered in one of the four quadrants were shown (quadrant localizer). These decreased in contrast according to the same spatial function that was used for the contrast modification in the pink noise stimuli. One localizer run consisted of the sequential presentation of localizer images for all quadrants. Each image was presented for 7.5s and changed polarity with 10Hz. The order of stimulation was: upper left, upper right, lower left and lower right and was repeated 10 times. All in all, the mapping and localization session consisted of eight different runs in the following order: 2x rotating wedge, expanding ring, 2x rotating wedge, expanding ring, 2x quadrant localizer. 155 volumes were acquired in each run, but no stimulation was present during the last 10s (five volumes). Participants had the same fixation task as in the pink noise experiment, with the only difference that the fixation spot changed every 1000ms.

Data processing

Functional brain scans were pre-processed with SPM2 (http://www.fil.ion.ucl.ac.uk/spm). The first five scans of each experimental run were discarded to allow for magnetic relaxation effects. All volumes acquired during one experimental session were motion corrected, realigned to the initial scan of the experiment, and coregistered to the high resolution anatomic image of the participant. For subsequent statistical analyses of the pink noise experiment, a general linear model (GLM) with event-based and HRF-convolved regressors was estimated separately for each voxel. For every run 9 regressors were used that encoded stimulation onsets: one regressor for no modification, four regressors for a high contrast modification in one of the quadrants and four regressors for a low contrast modification in one of the quadrants. In addition, a constant regressor for each run was included. All analyses were carried out based on SPM parameter estimates for these regressors.

Definition of functional regions of interest

Visual areas V1-V4 were functionally defined using well-established retinotopic mapping procedures (Wandell et al., 2007; Warnking, 2002). First, we segmented gray matter using FreeSurfer (Dale, Fischl, & Sereno, 1999). Next, the cortical surface was flattened with mrGray (Wandell, Chial, & Backus, 2000). Custom Matlab (The MathWorks, Natick, MA) scripts were used to generate the flattened angular phase maps (Heinzle, Kahnt, & Haynes, 2011). Finally, we identified visual areas V1-V4 by locating phase reversal boundaries on these maps. Figure 2 shows the outlines of V1-V4 on a flattened cortex for one participant and one hemisphere.

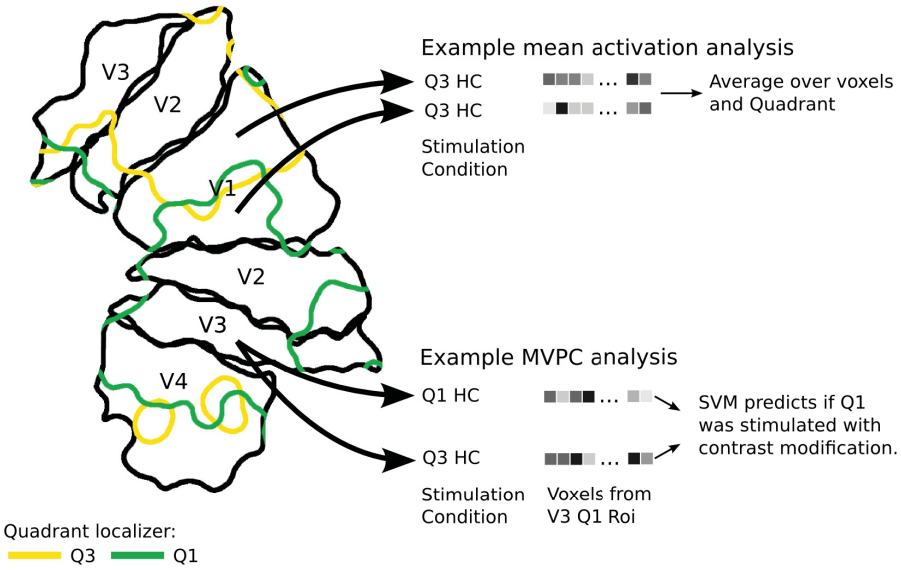

Figure 2. Depiction of mean activation analysis and the multivoxel pattern analysis. The black outlines show early visual areas identified by retinotopic mapping on a flattened cortex for one observer. The green outline marks areas that show significant activation upon stimulation of quadrant one with a flickering checkerboard. The yellow outline shows areas that are selective for stimulation of quadrant 3. The shaded green area highlights as an example the ROI V3 Q1. The two analyses were carried out for each visual area (mean activation analysis) and quadrant area combination (multivoxel pattern analysis).

The quadrants of the visual field were localized by fitting a GLM with one event-based and HRF-convolved regressor for each quadrant stimulation onset. Quadrant ROIs contained voxels that were exclusively active when one particular quadrant was stimulated (t-test $p<0.001$ uncorrected), but not during stimulation in one of the other three quadrants. Figure 2 shows outlines of quadrant specific regions for one participant and one hemisphere. Note that voxels which showed specificity for more than one quadrant were discarded.

Multivoxel pattern analysis (MVPA)

We trained support vector machines (SVM) for classifying whether a quadrant received a contrast modification. The SVMs were trained to predict, based on the activation of voxels in one ROI (e.g. V1 Q1), whether the quadrant corresponding to the ROI (Q1) or another quadrant was stimulated with modified contrast. This implicitly compares activation within a ROI when it is stimulated with a contrast modification with the activation when it is stimulated with baseline contrast. We conducted three separate classification analyses: First, a high contrast classifier that only received training and test data from conditions with a contrast increase in one quadrant; second, a low contrast classifier that only

received low contrast stimulation parameter estimates; third, a "saliency classifier" received data from both conditions (twice as many individual data points as the modification-specific classifiers), based on the rationale that neurons encoding saliency should show similar responses to both modifications.

SVM classification for each region of interest was based on three pairwise comparisons, separating contrast modification in the quadrant corresponding to the ROI (=stimulated) from modification in one of the other quadrants (=not stimulated). For example, the high contrast classification accuracy for the ROI V1 Q1 would be the mean of three accuracies: high contrast (hc) in Q1 vs. hc in Q2; hc in Q1 vs. hc in Q3; and hc in Q1 vs. hc in Q4. Training and evaluation of the SVMs was performed in a leave-one-out cross validation scheme. In each cross validation step, SPM model parameter estimates from 4 of the 5 experimental runs were used to train a classifier that predicted the location of the modified quadrant in the 5th run. This procedure was carried out for each participant and each ROI individually. The cost parameter was set to one for SVMs with a linear kernel. For RBF kernel SVMs a grid search was carried out for the cost and gamma parameter. The grid search was performed in every training step of the leave-one-out cross validation on training data only.

Results

High and low contrast modifications increase saliency

The primary goal of our study was to disentangle computations of luminance contrast and saliency. We created images on which luminance contrast in a quadrant was either decreased or increased. The attentional effect of these modifications was first investigated in an eye-tracking study. We recorded how often the first free fixation on an image fell into each quadrant. Analysis was restricted to the first fixation because its target is selected while the retinal stimulation is identical to the central fixation in the fMRI task. The distribution of fixations across the different quadrants is shown in Figure 3A. Each quadrant attracts more fixations in each of the two modification conditions than when it is unmodified (Fig. 3B). This is backed by a two factorial repeated measures analysis of variance with quadrant and modification as factors. Only the modification factor is significant (modification $p<0.001$; quadrant $p>0.3$; interaction $p>0.5$). Importantly, both modifications are significantly different from baseline (high contrast vs. baseline $p<0.001$; baseline vs. low contrast $p<0.001$, t-test). We conclude that both increases and decreases in local luminance contrast increase saliency in the modified quadrant by a comparable amount. Thus, these stimuli are suitable for disambiguating between the retinotopic processing of luminance contrast and saliency.

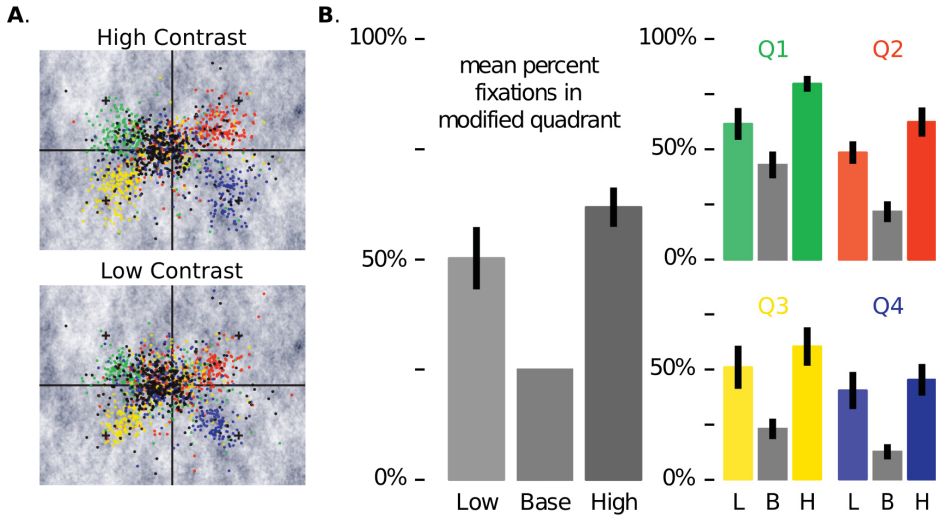

Figure 3. Distribution of fixations in the different stimulus conditions. A) Each color encodes fixations made when a certain quadrant was modified (Q1: green; Q2: red; Q3: yellow; Q4: blue). Gray fixations were made on unmodified stimuli. Solid gray lines mark the quadrant borders, plus-signs mark the peaks of the modification. Neither were shown on the actual stimuli. For all modifications, the fixation distribution is shifted towards the peak of the modification. B) Mean ratio of fixations made in each quadrant. Small colored figures show data for individual quadrants, the larger gray diagram shows the mean across all quadrants. Errorbars indicate the standard error of the mean across subjects. All quadrants attract more fixations when they are modified than in the baseline condition. This effect is independent of the direction of the contrast modification.

Mean BOLD activity increases with contrast

We analyzed how contrast modifications affected the mean BOLD response to the modified image regions in brain regions that process the visual input. We extracted GLM parameter estimates from all voxels in 16 functionally defined regions of interest (ROI) corresponding to the four quadrants of the visual field in V1-V4 (see method section and Figure 2 'Example mean activity analysis'). The contrast encoding hypothesis predicts low activity in quadrants stimulated with reduced contrast, and high activity for high contrast stimulation. The saliency encoding hypothesis, in its strongest form, predicts increased activity for both types of modification compared to baseline. We analyzed activity averaged across quadrants in V1-V4 in the high contrast condition, low contrast condition, and for the unmodified images (Figure 4). A repeated measures ANOVA with condition (high contrast, baseline, low contrast) and area (V1-V4) as factors reveals significant main effects of both factors ($p<0.001$), and a significant interaction ($p<0.05$). Single factor ANOVAs computed on the data of individual areas show that the effect of condition is significant throughout V1-V4 ($p<0.001$ Holm-Bonferroni corrected). We assessed the source of the significant effect with post hoc pairwise t-tests. The differences between high contrast and baseline as well as between high contrast and low contrast are significant in all areas ($p<0.01$). The difference between low contrast and baseline, the latter inducing higher activity than the former, is only significant in V1 ($p<0.01$) and V2 ($p<0.05$, all values Holm-Bonferroni corrected). In summary, high contrast leads to an increase in activity compared to both

baseline and low contrast condition, but low contrast, although salient, does not likewise lead to increased activity. To the contrary, if there is any difference between low contrast and baseline condition, it is not in the direction predicted by the saliency processing hypothesis.

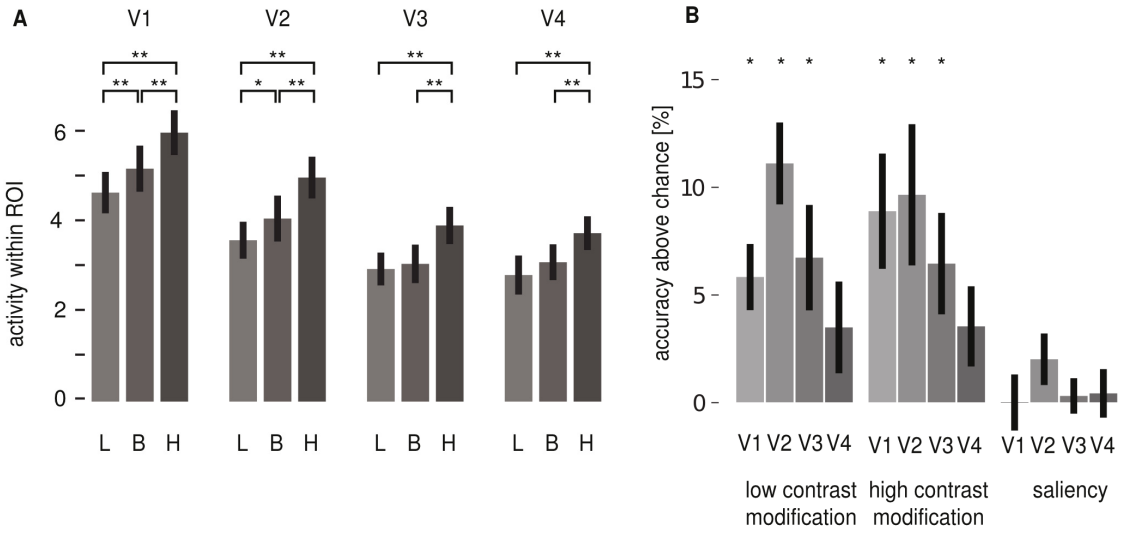

Figure 4. A) Mean BOLD activation in different visual areas in the 3 contrast conditions (L = low, B = baseline, H = High) averaged across quadrants. In all areas, an increase in contrast leads to either no change or an increase in BOLD signal, but never a decrease. Errorbars represent standard errors of the mean across subjects. Asterisks indicate significant differences between conditions (pairwise t-tests, Holm-Bonferroni corrected; *: $p<0.05$, **: $p<0.01$). B) Mean decoding accuracies above chance level for linear SVMs trained to predict whether a quadrant received a given modification. Errorbars represent *SEM* across subjects, asterisks indicate prediction performance significantly above chance assessed by a t-test ($p<0.05$).

MVPA supports contrast encoding hypothesis

In principle, it is possible that neurons in V1-V4 encode saliency, but that this information is represented in these areas in a way not accessible to an analysis of the activity level in the form of an averaged BOLD response. We used multivoxel pattern analysis (Haynes & Rees, 2006; Kriegeskorte, Goebel, & Bandettini, 2006) to test if information about the most salient quadrant can be decoded from the activity patterns in our ROIs. The contrast encoding hypothesis predicts that the stimulation of a visual field quadrant with a certain level of contrast leads to a specific pattern of activity in the ROI corresponding to that quadrant. It should thus be possible to decode whether the stimulus was modified in the quadrant corresponding to a ROI (see Figure 2). For example, given an activation pattern from ROI V3 Quadrant 1 (Q1), induced by high contrast stimulation in either Q1 or a different quadrant, it should be possible to infer if Q1 or another quadrant was modified (see Figure 2 'Example MVPC analysis'). The same should hold for low contrast modifications. The saliency encoding hypothesis furthermore predicts similar activation patterns for low and high contrast modifications, since both make a quadrant more salient (Figure 3). A classifier trained on both types of patterns combined should therefore be able to generalize, and infer if a quadrant was modified even without knowledge of the

modification type. Figure 5 shows the mean decoding accuracies achieved for the 3 different analyses (high contrast only, low contrast only, both contrasts mixed = saliency) in V1-V4. In areas V1 through V3, decoding accuracies were significantly above chance level for the high contrast only and low contrast only analyses (t-test across 12 subjects, $p<0.05$). However, the decoding accuracy for the saliency analysis did not reach significance in any ROI. Since the difference between a significant result and a non-significant one is not necessarily itself significant (Gelman & Stern, 2006; Nieuwenhuis, Forstmann, & Wagenmakers, 2011), we also directly analyzed the differences in accuracy between the contrast classifiers and the saliency classifier. Here, we find that in areas V1 through V3, decoding accuracy is significantly higher for the contrast classifiers. Results for V4 are not significant, but the trend goes in the direction predicted by the contrast encoding hypothesis.

To ensure that the failure to decode saliency from early visual areas is not due to specific parameters chosen, we performed additional analyses. First, we used SVMs with an RBF kernel instead of a linear one. Second, we trained SVMs on activation patterns from whole areas, instead of only single quadrants. And third, we used anatomical ROI definitions from the SPM anatomy toolbox (Amunts, Malikovic, Mohlberg, Schormann, & Zilles, 2000; Eickhoff et al., 2005; Rottschy et al., 2007) instead of the functional ones. Neither of these changes, nor combinations thereof, affected the pattern of results reported for the linear SVMs.

In summary, it is possible to decode whether a quadrant was modified when modification type is given. However, without this information it is not possible to decode whether a quadrant is salient. This suggests that V1-V4 do not make the abstraction away from absolute changes in contrast to changes in saliency.

Discussion

We showed that low and high contrast modifications in pink noise stimuli decouple saliency and contrast. Our eye tracking data indicate that both types of modification increase saliency. This decoupling provides a tool for investigating saliency processing in fMRI BOLD responses in early topographically organized visual areas (V1-V4). The behavioral increase in saliency for the low contrast modifications is not mirrored in fMRI data. Instead, we found that the activity patterns of these regions monotonically relate to stimulus contrast, not saliency.

In order to encode saliency for these stimuli, the visual system would have to increase its response to contrast deviations from the mean in both directions. Gardner et al. (2005) have shown that such a rectification operation may happen in V4 during temporal contrast adaptation. It might have been suspected that a similar mechanism for spatial variations in contrast is responsible for the behavioral saliency effect observed in our stimuli. We do not find evidence for this, i.e. saliency could not be decoded in V4. However, contrast could also not be decoded in V4, which might be indicative of a low

signal to noise ratio. The question of whether saliency is encoded in V4 can therefore not be conclusively addressed with our data.

The "V1 saliency hypothesis" (Li, 2002) states that activity in V1 creates a bottom-up saliency map. Specifically, the highest evoked V1 response of each visual field location (i.e. a max operation over all features encoded for this location) gives the relative saliency of this location. There is psychophysical (Zhaoping & May, 2007) as well as physiological (Zhang et al., 2012) evidence supporting this hypothesis. At first sight these data appear to be in conflict with the present results. However, their stimuli are not natural stimuli, but arrays of oriented bars or simple conjunctions of bars. It is known that the receptive fields of V1 neurons are highly tuned to such bars. Under these conditions it is therefore plausible that processing in V1 contributes to a saliency map. Given the restricted stimulus set, focusing on oriented line elements, the intermediate results might be indistinguishable from the final saliency map. In contrast, we used stimuli with a power spectrum that is comparable to natural scenes. Recent work demonstrates that such stimuli induce qualitatively different dynamics in visual cortex than gratings (Onat, König, & Jancke, 2011). Hence, our more complex stimuli might explain why we find that V1 BOLD activity only contributes one processing step on the way to a final saliency map. This is consistent with recent results on experimental blindsight in monkeys (Yoshida et al., 2012). Interestingly, there is even evidence for salient orientation pop-out stimuli which are represented in V4 rather than in V1 (Bogler, Bode, & Haynes, 2013). These results are not compatible with the predictions of a general saliency map localized in primary visual cortex.

It should be noted that our results do not rule out contributions of V1-V4 to the computation of saliency even in the low-contrast modification condition. It might, for example, be that subpopulations of neurons in these areas compute saliency and that the activity of these subpopulations is swamped by the contrast dependent activity changes of the majority of neurons. However, (Zhang et al., 2012) do find an explicit attention driven signal in BOLD responses in V1 regions even for stimuli that were not consciously perceived. Thus, it does not seem that the proposed V1 saliency map is in principle not discoverable with fMRI. This concern is further reduced by our use of decoding techniques. It has been shown that MVPC analyses can be successfully used to decode the activity of neuronal subpopulations below the spatial resolution of individual fMRI voxels (Haynes & Rees, 2005). But this is of course still no guarantee that decoding would have been successful in our case if saliency is encoded by a small set of neurons in early visual areas. However, more explicit representations of saliency are observable in higher visual areas with fMRI (Bogler, Bode, & Haynes, 2011). Summarizing, the most dominant feature in V1-V4, according to our analysis, is clearly luminance contrast and not saliency.

In conclusion, we report a case of behaviorally observable saliency that is not linearly driven by stimulus contrast. Our findings do not support the hypothesis that a saliency map, in the sense of an explicit representation of most likely fixation target regardless of specific stimulus features, is found in V1-V4. It is conceivable that higher areas have to integrate feature specific saliency information encoded in early processing stages.

Acknowledgements

This work was supported by two grants by the German Federal Ministry of Education and Research (BMBF grant 01GQ1001C, BMBF grant 01GQ0851), by the EU through the project eSMCs (FP7-IST-270212) and ERC-2010-AdG #269716 - MULTISENSE, the Deutsche Forschungsgemeinschaft (GRK1589/1), and the Max Planck Society.

References

Açik, A., Onat, S., Schumann, F., Einhäuser, W., & König, P. (2009). Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories. Vision research, 49(12), 1541–53. doi:10.1016/j.visres.2009.03.011

Açık, A., Sarwary, A., Schultze-Kraft, R., Onat, S., & König, P. (2010). Developmental Changes in Natural Viewing Behavior: Bottom-Up and Top-Down Differences between Children, Young Adults and Older Adults. Frontiers in psychology, 1(November), 207. doi:10.3389/fpsyg.2010.00207

Bogler, C., Bode, S., & Haynes, J.-D. (2011). Decoding successive computational stages of saliency processing. Current biology : CB, 21(19), 1667–71. doi:10.1016/j.cub.2011.08.039

Bogler, C., Bode, S., & Haynes, J.-D. (2013). Orientation pop-out processing in human visual cortex. NeuroImage, 81, 73–80. doi:10.1016/j.neuroimage.2013.05.040

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. Annual review of neuroscience, 33, 1–21. doi:10.1146/annurev-neuro-060909-152823

Dale, A., Fischl, B., & Sereno, M. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. Neuroimage, 194, 179–194. Retrieved from http://www.sciencedirect.com/science/article/pii/S1053811998903950

Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? European Journal of Neuroscience, 17(5), 1089–1097. doi:10.1046/j.1460-9568.2003.02508.x

Einhäuser, Wolfgang, Rutishauser, U., Frady, E. P., Nadler, S., König, P., & Koch, C. (2006). The relation of phase noise and luminance contrast to overt attention in complex visual stimuli. Journal of vision, 6(11), 1148–58. doi:10.1167/6.11.1

Gardner, J. L., Sun, P., Waggoner, R. A., Ueno, K., Tanaka, K., & Cheng, K. (2005). Contrast adaptation and representation in human early visual cortex. Neuron, 47(4), 607–20. doi:10.1016/j.neuron.2005.07.016

Gelman, A., & Stern, H. (2006). The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. The American Statistician, 60(4), 328–331. doi:10.1198/000313006X152649

Geng, J. J., & Mangun, G. R. (2009). Anterior intraparietal sulcus is sensitive to bottom-up attention driven by stimulus salience. Journal of cognitive neuroscience, 21(8), 1584–601. doi:10.1162/jocn.2009.21103

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. Nature, 391(6666), 481–4. doi:10.1038/35135

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. Nature

reviews. Neuroscience, 7(7), 523–34. doi:10.1038/nrn1931

Heinzle, J., Kahnt, T., & Haynes, J.-D. (2011). Topographically specific functional connectivity between visual field maps in the human brain. NeuroImage, 56(3), 1426–36. doi:10.1016/j.neuroimage.2011.02.077

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. The Journal of physiology, 148(3), 574-591.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. Nat Rev Neurosci, 2(3), 194–203.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. Human neurobiology. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3836989

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. Proceedings of the National Academy of Sciences of the United States of America, 103(10), 3863–8. doi:10.1073/pnas.0600244103

Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. Proceedings of the National Academy of Sciences of the United States of America, 96(18), 10530–5. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=17923&tool=pmcentrez&rendertype=abstract

Li, Zhaoping. (2002). A saliency map in primary visual cortex. Trends in Cog. Sci.

Mazer, J. a, & Gallant, J. L. (2003). Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. Neuron, 40(6), 1241–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14687556

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. Nature Neuroscience, 14(9), 1105–1107. doi:10.1038/nn.2886

Onat, S., König, P., & Jancke, D. (2011). Natural scene evoked population dynamics across cat primary visual cortex captured with voltage-sensitive dye imaging. Cerebral cortex (New York, N.Y. : 1991), 21(11), 2542–54. doi:10.1093/cercor/bhr038

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. Psychological science, 16(2), 114–22. doi:10.1111/j.0956-7976.2005.00791.x

Serences, J. T., & Yantis, S. (2007). Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. Cerebral cortex (New York, N.Y. : 1991), 17(2), 284–93. doi:10.1093/cercor/bhj146

Shipp, S. (2004). The brain circuitry of attention. Trends in Cognitive Sciences, 8(5), 223–230.

Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. Progress in brain research, 249–262. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0079612304470198

Wandell, B., Chial, S., & Backus, B. T. (2000). Visualization and measurement of the cortical surface. Journal of cognitive neuroscience, 12(5), 739–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11054917

Wandell, B., Dumoulin, S. O., & Brewer, A. (2007). Visual field maps in human cortex. Neuron, 56(2), 366–83. doi:10.1016/j.neuron.2007.10.012

Warnking, J. (2002). fMRI Retinotopic Mapping—Step by Step. NeuroImage, 17(4), 1665–1683.

doi:10.1006/nimg.2002.1304

Yoshida, M., Itti, L., Berg, D. J., Ikeda, T., Kato, R., Takaura, K., White, B. J., Munoz, D. P., & Isa, T. (2012). Residual attention guidance in blindsight monkeys watching complex natural scenes. Current Biology 22(15), 1429-1434. doi: 10.1016/j.cub.2012.05.046

Zhang, X., Zhaoping, L., Zhou, T., & Fang, F. (2012). Neural activities in v1 create a bottom-up saliency map. Neuron, 73(1), 183–92. doi:10.1016/j.neuron.2011.10.035

Zhaoping, L. (2011). Neural circuit models for computations in early visual cortex. Current opinion in neurobiology, 21(5), 808–15. doi:10.1016/j.conb.2011.07.005

Zhaoping, L., & May, K. a. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. PLoS computational biology, 3(4), e62. doi:10.1371/journal.pcbi.0030062

6.3 Decoding Successive Computational Stages of Saliency Processing

Bogler C., Bode S., Haynes J.D. (2011). Decoding successive computational stages of saliency processing. *Curr Biol*, 21(19):1667-71.

Current Biology 21, 1667–1671, October 11, 2011 ©2011 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2011.08.039

Report

Decoding Successive Computational Stages of Saliency Processing

Carsten Bogler,[1,2,3,*] Stefan Bode,[2,3] and John-Dylan Haynes[1,2,3,*]
[1]Bernstein Center for Computational Neuroscience Berlin, Charité – Universitätsmedizin Berlin, 10115 Berlin, Germany
[2]Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany
[3]Department of Neurology, Otto-von-Guericke University, 39120 Magdeburg, Germany

Summary

An important requirement for vision is to identify interesting and relevant regions of the environment for further processing. Some models assume that salient locations from a visual scene are encoded in a dedicated spatial saliency map [1, 2]. Then, a winner-take-all (WTA) mechanism [1, 2] is often believed to threshold the graded saliency representation and identify the most salient position in the visual field. Here we aimed to assess whether neural representations of graded saliency and the subsequent WTA mechanism can be dissociated. We presented images of natural scenes while subjects were in a scanner performing a demanding fixation task, and thus their attention was directed away. Signals in early visual cortex and posterior intraparietal sulcus (IPS) correlated with graded saliency as defined by a computational saliency model. Multivariate pattern classification [3, 4] revealed that the most salient position in the visual field was encoded in anterior IPS and frontal eye fields (FEF), thus reflecting a potential WTA stage. Our results thus confirm that graded saliency and WTA-thresholded saliency are encoded in distinct neural structures. This could provide the neural representation required for rapid and automatic orientation toward salient events in natural environments.

Results

An object in a visual scene that is different than its surround automatically captures one's attention or pops out. This could be, for example, a man wearing a yellow suit or just a horizontally oriented bar among vertical ones. The visual system is automatically guided to process such salient objects because they are believed to be most informative and relevant. Each item in a visual scene can be thought to have a "saliency," specifying its relative quality to stand out among the other items. It has often been proposed that saliency information is represented in a spatial map [1, 2] that encodes the saliency for every position in the visual field, although this has been debated [5]. Because saliency is based on low-level sensory features, it is referred to as "bottom-up" attentional control. Besides such bottom-up effects, attention can also be controlled in a "top-down" fashion based on memory or behavioral goals.

According to a prominent model by Itti and Koch [1, 2], a local feature gradient is first computed separately for different feature dimensions (such as color, orientation, or luminance) and then integrated to an overall saliency value. This model does not explicitly specify the neural implementation of the saliency map in the brain. Various locations of a saliency map have been proposed, including subcortical structures such as superior colliculus (SC) [6] and pulvinar [7], primary visual cortex (V1) [8, 9], the ventral visual pathway [5, 10, 11], the intraparietal sulcus (IPS) [11, 12], the human homolog of the lateral intraparietal area (LIP) of the monkey, and the frontal eye fields (FEF) [11, 13]. Importantly, to date it has remained unclear where the transition would occur between graded saliency signals and a winner-take-all (WTA)-thresholded representation of the maximally salient position.

Here, we aimed to disentangle the different stages of saliency processing: (1) the graded representation of saliency for four quadrants and (2) the winner-take-all-thresholded representation of the maximally salient position in the visual field. We presented our subjects with natural scenes so that saliency could be based on multiple, naturalistic low-level features (Figure 1). Natural scenes also have image statistics to which the visual system is tuned to and that therefore are optimal for automatic processing [14]. Then we used a computational model [2, 15] to estimate the saliency at each location in the visual field and averaged the saliency in four separate sectors (Figure 2).

We used whole-brain functional imaging in combination with a general linear model and multivariate pattern classification methods to search for saliency-related information (see the Supplemental Information available online for detailed experimental procedures). First, a parametric general linear model (GLM) analysis was used to identify brain regions in which the blood oxygen level-dependent (BOLD) signal increased linearly with a gradual increase of saliency in the images. This analysis revealed that neural activity in bilateral striate and extrastriate cortex as well as left IPS was correlated with the graded saliency of the images (Figure 3, red regions). Importantly, only saliency signals in early visual cortex could be traced to individual quadrants (Figure 4), which would be expected from their retinotopic structure. In contrast, graded saliency representations in parietal cortex reflected the superposition of multiple quadrants, possibly due to the lack of resolution of functional magnetic resonance imaging (fMRI).

Then we used multivariate pattern classification to identify areas that encode the thresholded output reflecting the most salient quadrant. For this we used a searchlight classification approach [3, 4] that assesses in an unbiased fashion to which degree the thresholded saliency can be decoded from the local cluster of voxels at each position in the brain. The maximally salient locations in the images (which were considered to reflect the outcome of a hypothetical WTA mechanism), could be successfully decoded from left and right IPS (both $p < 0.05$, family-wise error [FWE] corrected; peak accuracies: 33.6% and 32.1% in left and right IPS, correspondingly; chance level: 25%; Figure 3, blue regions). Interestingly, the region in the IPS that showed strong correlation with the initial saliency map and the region that encoded the most salient position showed no overlap. The region in the IPS found by multivariate pattern classification to encode WTA-thresholded saliency

*Correspondence: carsten.bogler@bccn-berlin.de (C.B.), haynes@bccn-berlin.de (J.-D.H.)

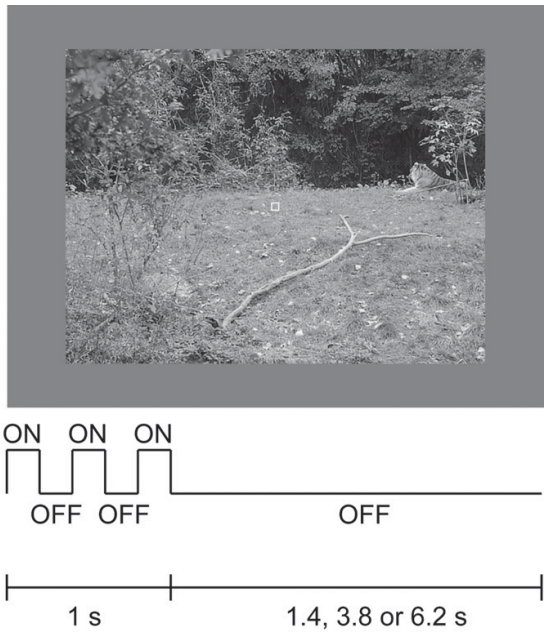

Figure 1. Visual Stimulation

In each trial, one image was presented repeatedly for 200 ms with a 200 ms gap. Between successive trials there was a variable interstimulus interval of 1.4 to 6.2 s. Subjects fixated on a demanding task at the center of the screen. Every 1,200 ms, the left or right bar of the square at the center of the screen was removed, and subjects had to indicate whether the square opened up to the left- or right-hand side.

was located more anterior than the regions revealed by parametric analysis of graded saliency (Figure 3). Additional regions encoding the most salient position were the bilateral FEF ($p < 0.05$, FWE corrected; peak accuracies: 35% and 34.2% in left and right FEF, correspondingly).

Support vector machines classify two different classes. Multiclass classification is typically realized by combining multiple pairwise classifications. Thus accuracy above chance level in a four-class pattern classification could theoretically be the result of perfect or close to perfect classification of only a subset of the classes. Thus, we further clarified that our analysis reflects information from all quadrants. Therefore, we ran all six possible pairwise searchlight multivariate pattern classification analyses. The results confirmed that it was possible to decode the most salient quadrant of any two quadrants from right and left IPS and FEF (see Figure S2). The only exception was the right FEF from which it was not possible to decode one combination (lower right versus upper right quadrant; note that chance level was 50% for this analysis).

Discussion

By presenting our subjects with photographs of natural scenes in a rapid event-related fMRI experiment, we identified brain regions associated with different stages of a bottom-up attention model [1]. Activation levels in visual cortex and left posterior IPS (pIPS) correlated with the graded saliency in different parts of the photographs. Using multivariate pattern classification, we could further demonstrate that bilateral anterior IPS and the FEF encoded information about the maximally salient quadrant, thus possibly reflecting the outcome of a WTA mechanism.

The correlation between the saliency of the four quadrants and the BOLD response in visual cortex and pIPS suggests that these areas are involved in calculating saliency information. Zhaoping [9] previously argued for a saliency map representation in V1, mainly based on psychophysical experiments and theoretical considerations about the V1 architecture. Some studies [16, 17] support the V1 saliency map hypothesis, and our finding of a graded saliency representation in early visual cortex is also compatible with this model. On the other hand, Hegdé and Felleman [18] reported that V1 neurons generally responded to feature discontinuities that do not necessarily have to be salient. Furthermore, regions of extrastriate visual cortex (V4) respond to pop-out stimuli [10, 19], however, possibly only if attention is directed to the relevant feature [19].

In line with our data, different regions within the parietal cortex were identified to show a direct saliency representation in fMRI studies with humans [20] as well as in physiological studies with monkeys [21–24]. One hypothesis is that conspicuity in elementary feature contrast maps is encoded in visual cortex, whereas saliency (integrated across multiple dimensions) is encoded in pIPS. This reasoning is compatible with the saliency map model [1], where feature contrast maps for luminance, orientation, and color are calculated first and then combined to the saliency map. Computationally, this

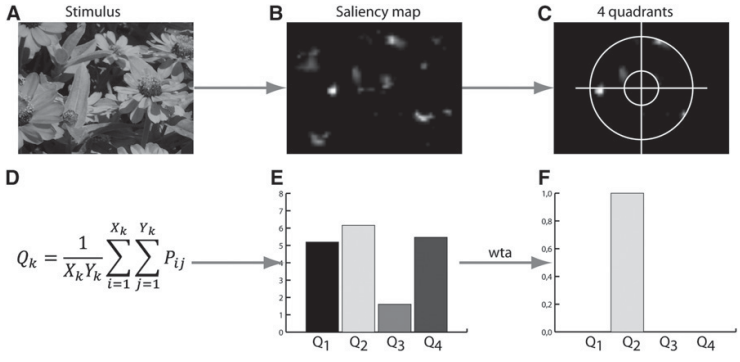

Figure 2. Extracting the Saliency Representation for Natural Visual Stimuli

(A) One hundred different images of natural scenes were presented during the experiment.

(B) For each image, the saliency map was calculated based on an implementation of Itti and Koch's saliency map model [2, 15].

(C) The saliency was then averaged across individual four sectors, defined by screening central and peripheral regions out of each visual field quadrant.

(D–F) The average saliency for each quadrant (D) was then used to define four parameters for each image, which encode the graded saliency (E), and a winner-take-all (WTA) mechanism thresholded the four saliency values so that only the most salient quadrant remained (F).

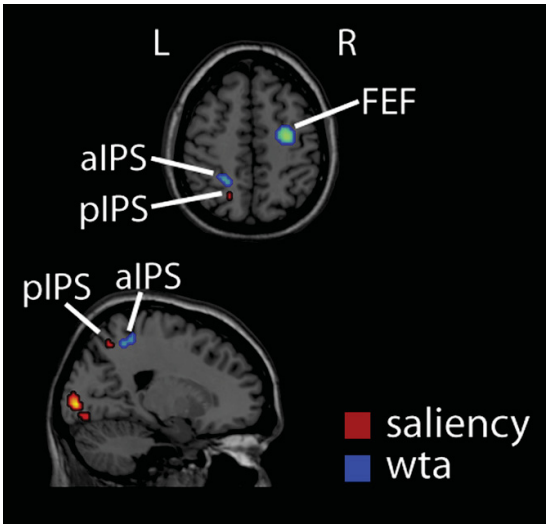

Figure 3. Two Stages of Saliency Representation

Regions in visual cortex and posterior intraparietal sulcus (pIPS) that correlated with the graded saliency map (red) and regions in the anterior IPS and frontal eye fields (FEF) that encoded the output of the WTA stage, i.e., the most salient quadrant (blue). In IPS, the WTA type of code could be found more anterior compared to the graded saliency representation (both $p < 0.05$, family-wise error corrected).

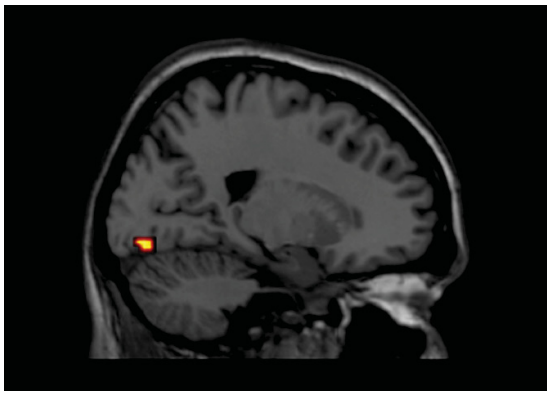

Figure 4. Quadrant-Specific Saliency

Among the regions that correlated with graded saliency (Figure 3), only visual cortex responded in a selective fashion to the saliency in only one of the quadrants (F test $p < 0.0005$, minimum cluster size ten voxels).

would be realized by averaging the individual feature contrast maps. It has been shown that neurons in visual cortex have the appropriate properties for the calculation of elementary feature contrast maps. Additionally, the feed-forward connections from visual areas to IPS provide a potential anatomical substrate for the integration of feature gradients across different dimensions. Alternatively, saliency could be computed progressively in several successive stages [25].

In many studies that investigated the IPS response to bottom-up saliency, modulation by top-down attentional factors could also be demonstrated. However, similar attentional top-down modulation was found for visual areas including lateral geniculate nucleus (LGN) [26], V1 [27], V4 [27], and MT+/V5 [28], and this does not contradict a potential saliency map representation in IPS.

Please note that our study was not designed to conclusively reveal the exact topography of a saliency map. Nevertheless, in visual cortex, but not in IPS, we found voxels that showed a higher response to the saliency in only one of the quadrants relative to the other three quadrants (Figure 4), thus suggesting a retinotopic representation of saliency in visual cortex. In pIPS, however, the BOLD response in each voxel was informative of the saliency in two or more quadrants. There are several possible explanations for this finding: First, the size of IPS is smaller than that of V1, which means that it will have been sampled by fewer fMRI voxels, thus potentially obscuring any retinotopic structure [29]. Second, IPS neurons have larger receptive fields [30], which make it difficult to identify retinotopic organization. Third, the high anatomical variability of the IPS across subjects might have obscured any retinotopic effects in the averaged data. However, please note that the quadrant-based analysis used here does not require the full topography of the maps to be identified.

Low-level stimulus properties that are used to calculate the saliency map contribute to human overt attention [31]. However, it has been shown that low-level stimulus features correlate with high-level features [32–34]. Therefore, it might be possible that part of the decoded information about the most salient quadrant is related to high-level concepts. A study that investigated saccades of patients with visual object agnosia contradicts this possibility [35]. It could be shown that the first saccades seem to be controlled by low-level features in contrast to the late saccades, which seem to be controlled by high-level features.

A successful attention shift requires not only information about the graded saliency but also about the most salient position in the visual field, i.e., the output of the WTA-stage (see Figure S4 for a discussion of explicit and implicit representations). Here we were able to identify two regions, the anterior intraparietal sulcus (aIPS) and the FEF, that clearly match the output of a WTA-based model. Using multivariate pattern classification, information about the most salient quadrant could be found in bilateral aIPS and bilateral FEF. One explanation for this is that the saliency representation in pIPS is thresholded by a WTA mechanism and that the information about the most salient quadrant is then propagated to aIPS and FEF. This means that the WTA computation is localizable rather than being an emergent property of recurrent processing in multiple spatial maps [36]. Multivariate pattern classification using a searchlight approach was performed with a fixed size of the searchlight for the whole brain. Receptive field size of neurons increases from lower to higher visual areas. This might bias the results in favor of higher brain areas like the aIPS and the FEF as to encode information about WTA-thresholded saliency because a larger part of the represented visual field can be covered in these regions. However, note that a multivariate pattern regression using a similar searchlight approach (see Figure S1) was used to search for information about graded saliency. Despite any potential biases to favor higher brain areas, only visual cortex was identified to encode distributed information about graded saliency. Furthermore, we compared decoding of WTA saliency from our searchlight to decoding from a large V1 region of interest (see Figure S3) in order to ensure that our finding was not

due to the restricted spatial extent of the searchlight classifier. The comparison confirmed that there was no difference in decoding accuracy between both approaches (see Supplemental Information). This gives evidence that indeed different information is encoded in visual cortex and aIPS and FEF and that the result is not due to a bias that relies on the fixed size of the searchlight.

The IPS has been associated with saliency representation in previous studies [20, 21, 23], but the distinction between coding of saliency and coding of the output of a WTA stage was typically not clarified. Similarly, the FEF was reported to encode a saliency or priority map [13], which is defined as a saliency map with strong behavioral relevant top-down influence [37]. Also, in primates, FEF has been shown to encode saliency signals even when they pertain to objects that are not the goal of a current search task [13]. It has also been shown with fMRI that FEF responds to stimulus-driven attention [38], but again, it was not clear in these studies whether this stage reflected the graded coding of saliency or the WTA stage. Previously, the IPS was found to be sensitive to bottom-up attention, in contrast to the FEF, which was found to be involved in top-down attention [20].

Interestingly, even though attention was directed away from the stimulus, the regions we identified as reflecting the WTA-stage overlap with regions previously reported to be involved in control of overt and covert attention [39, 40]. Corbetta and Shulman [41] distinguish two neuronal attentional networks: the ventral frontoparietal and the dorsal frontoparietal networks. The dorsal frontoparietal network is supposed to guide top-down or goal-driven attention, whereas the ventral frontoparietal network should enable the detection of salient stimuli. These areas were also identified during saccadic eye movements [23, 29, 39]. Additionally, previous studies [39, 40] could demonstrate that the same cortical networks are active during overt and covert shifts of attention. Thus shifts of covert or overt attention could potentially cause a similar result pattern. However, please note that our study was designed to minimize the effects of shifting attention on the encoding of saliency. In order to direct attention away from salient locations of the stimuli, our subjects were engaged in a demanding fixation task. The most salient quadrant was not behaviorally relevant, thus giving subjects no reason to initiate saccaded to this location. In an additional inattentional blindness experiment (see Supplemental Information), we could also demonstrate that subjects were not able to indicate the most salient quadrant while they were performing this demanding fixation task. In a previous study, it could be demonstrated that a very similar fixation task to the one used here reduced the hemodynamic response in visual cortex for task-irrelevant images [42]. It is crucial to note that the purpose of final WTA computations is to prepare potential shifts of attention to interesting, salient positions. We were therefore able to successfully isolate the cascade of automatic, saliency-based orientation preparation from the actual overt action, even under conditions in which attention was bound to fixation. Thus, our results suggest that the WTA operation takes place automatically and does not require attention. There have been several demonstrations of neural processing of various features, including saliency, in the absence of task-relevance and attention (e.g., [42, 43]).

Taken together, our results support a computational bottom-up saliency model and furthermore associate different anatomical regions to different computational stages of the model. Graded saliency is represented in visual cortex and pIPS. Information about the most salient position is finally extracted from the graded saliency representations yielding a representation of the most salient quadrant (WTA mechanism) in aIPS and FEF. This signal might be related to performing shifts of attention. Methods highly sensitive for fine-grained local information, such as multivariate pattern classification, could identify automatic, unconscious preparation for orientation, although subjects actually did not overtly or covertly shift their attention.

Supplemental Information

Supplemental Information includes four figures, three tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.cub.2011.08.039.

Acknowledgments

We would like to thank Felix Wichmann for providing the stimuli. This work was funded by the Bernstein Computational Neuroscience Program of the German Federal Ministry of Education and Research (BMBF grant 01GQ0411), the German Research Foundation (DFG HA 5336/1-1), the Excellence Initiative of the German Federal Ministry of Education and Research (DFG grant GSC86/1-2009), and the Max Planck Society.

Received: February 17, 2011
Revised: July 22, 2011
Accepted: August 16, 2011
Published online: September 29, 2011

References

1. Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol. *4*, 219–227.
2. Itti, L., and Koch, C. (2001). Computational modelling of visual attention. Nat. Rev. Neurosci. *2*, 194–203.
3. Haynes, J.D., and Rees, G. (2006). Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. *7*, 523–534.
4. Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. Proc. Natl. Acad. Sci. USA *103*, 3863–3868.
5. VanRullen, R. (2003). Visual saliency and spike timing in the ventral visual pathway. J. Physiol. Paris *97*, 365–377.
6. Kustov, A.A., and Robinson, D.L. (1996). Shared neural control of attentional shifts and eye movements. Nature *384*, 74–77.
7. Robinson, D.L., and Petersen, S.E. (1992). The pulvinar and visual salience. Trends Neurosci. *15*, 127–132.
8. Gao, D., Mahadevan, V., and Vasconcelos, N. (2008). On the plausibility of the discriminant center-surround hypothesis for visual saliency. J. Vis. *8*, 1–18.
9. Li, Z. (2002). A saliency map in primary visual cortex. Trends Cogn. Sci. (Regul. Ed.) *6*, 9–16.
10. Mazer, J.A., and Gallant, J.L. (2003). Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. Neuron *40*, 1241–1250.
11. Serences, J.T., and Yantis, S. (2007). Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. Cereb. Cortex *17*, 284–293.
12. Gottlieb, J.P., Kusunoki, M., and Goldberg, M.E. (1998). The representation of visual salience in monkey parietal cortex. Nature *391*, 481–484.
13. Thompson, K.G., and Bichot, N.P. (2005). A visual salience map in the primate frontal eye field. Prog. Brain Res. *147*, 251–262.
14. Einhäuser, W., and König, P. (2010). Getting real-sensory processing of natural stimuli. Curr. Opin. Neurobiol. *20*, 389–395.
15. Walther, D., and Koch, C. (2006). Modeling attention to salient proto-objects. Neural Netw. *19*, 1395–1407.
16. Kastner, S., Nothdurft, H.C., and Pigarev, I.N. (1997). Neuronal correlates of pop-out in cat striate cortex. Vision Res. *37*, 371–376.
17. Albright, T.D., and Stoner, G.R. (2002). Contextual influences on visual processing. Annu. Rev. Neurosci. *25*, 339–379.
18. Hegdé, J., and Felleman, D.J. (2003). How selective are V1 cells for pop-out stimuli? J. Neurosci. *23*, 9968–9980.

19. Burrows, B.E., and Moore, T. (2009). Influence and limitations of popout in the selection of salient visual stimuli by area V4 neurons. J. Neurosci. *29*, 15169–15177.
20. Geng, J.J., and Mangun, G.R. (2009). Anterior intraparietal sulcus is sensitive to bottom-up attention driven by stimulus salience. J. Cogn. Neurosci. *21*, 1584–1601.
21. Gottlieb, J. (2007). From thought to action: the parietal cortex as a bridge between perception, action, and cognition. Neuron *53*, 9–16.
22. Bisley, J.W., and Goldberg, M.E. (2006). Neural correlates of attention and distractibility in the lateral intraparietal area. J. Neurophysiol. *95*, 1696–1717.
23. Goldberg, M.E., Bisley, J.W., Powell, K.D., and Gottlieb, J. (2006). Saccades, salience and attention: the role of the lateral intraparietal area in visual behavior. Prog. Brain Res. *155*, 157–175.
24. Constantinidis, C., and Steinmetz, M.A. (2005). Posterior parietal cortex automatically encodes the location of salient stimuli. J. Neurosci. *25*, 233–238.
25. Soltani, A., and Koch, C. (2010). Visual saliency computations: mechanisms, constraints, and the effect of feedback. J. Neurosci. *30*, 12831–12843.
26. O'Connor, D.H., Fukui, M.M., Pinsk, M.A., and Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. Nat. Neurosci. *5*, 1203–1209.
27. McAdams, C.J., and Maunsell, J.H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J. Neurosci. *19*, 431–441.
28. Martinez-Trujillo, J.C., and Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr. Biol. *14*, 744–751.
29. Sereno, M.I., Pitzalis, S., and Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. Science *294*, 1350–1354.
30. Ben Hamed, S., Duhamel, J.R., Bremmer, F., and Graf, W. (2001). Representation of the visual field in the lateral intraparietal area of macaque monkeys: a quantitative receptive field analysis. Exp. Brain Res. *140*, 127–144.
31. Kollmorgen, S., Nortmann, N., Schröder, S., and König, P. (2010). Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. PLoS Comput. Biol. *6*, e1000791.
32. Einhäuser, W., and König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? Eur. J. Neurosci. *17*, 1089–1097.
33. Einhäuser, W., Spain, M., and Perona, P. (2008). Objects predict fixations better than early saliency. J. Vis. *8*, 1–26.
34. Elazary, L., and Itti, L. (2008). Interesting objects are visually salient. J. Vis. *8*, 1–15.
35. Mannan, S.K., Kennard, C., and Husain, M. (2009). The role of visual salience in directing eye movements in visual object agnosia. Curr. Biol. *19*, R247–R248.
36. Bruce, N.D.B., and Tsotsos, J.K. (2009). Saliency, attention, and visual search: an information theoretic approach. J. Vis. *9*, 5, 1–24.
37. Fecteau, J.H., and Munoz, D.P. (2006). Salience, relevance, and firing: a priority map for target selection. Trends Cogn. Sci. (Regul. Ed.) *10*, 382–390.
38. Kincade, J.M., Abrams, R.A., Astafiev, S.V., Shulman, G.L., and Corbetta, M. (2005). An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. J. Neurosci. *25*, 4593–4604.
39. Corbetta, M., Akbudak, E., Conturo, T.E., Snyder, A.Z., Ollinger, J.M., Drury, H.A., Linenweber, M.R., Petersen, S.E., Raichle, M.E., Van Essen, D.C., and Shulman, G.L. (1998). A common network of functional areas for attention and eye movements. Neuron *21*, 761–773.
40. Perry, R.J., and Zeki, S. (2000). The neurology of saccades and covert shifts in spatial attention: an event-related fMRI study. Brain *123*, 2273–2288.
41. Corbetta, M., and Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. *3*, 201–215.
42. Tusche, A., Bode, S., and Haynes, J.D. (2010). Neural responses to unattended products predict later consumer choices. J. Neurosci. *30*, 8024–8031.
43. Bichot, N.P., Chenchal Rao, S., and Schall, J.D. (2001). Continuous processing in macaque frontal cortex during visual search. Neuropsychologia *39*, 972–982.

Current Biology, Volume *21*

Supplemental Information

Decoding Successive Computational Stages of Saliency Processing

Carsten Bogler, Stefan Bode, and John-Dylan Haynes

Supplemental Inventory

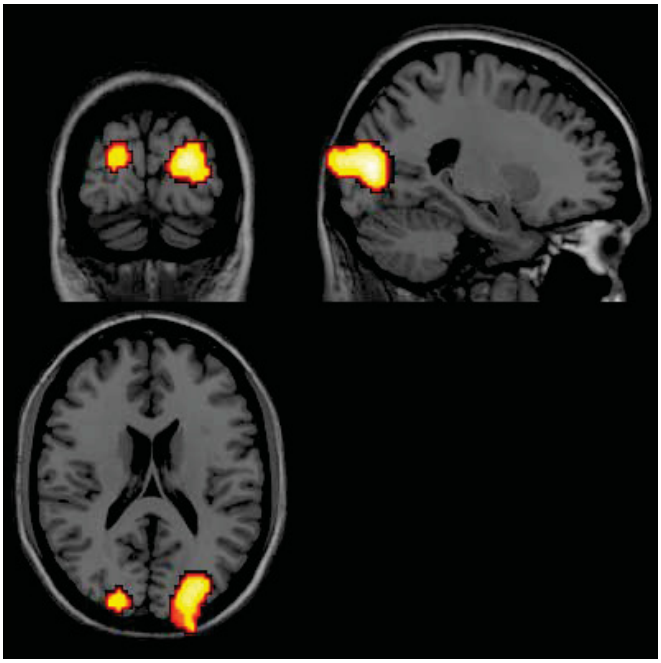

Figure S1.
In visual cortex graded saliency could be decoded from distributed patterns using multivariate support vector regression ($p < 0.0001$ uncorrected, cluster size >10 voxels).

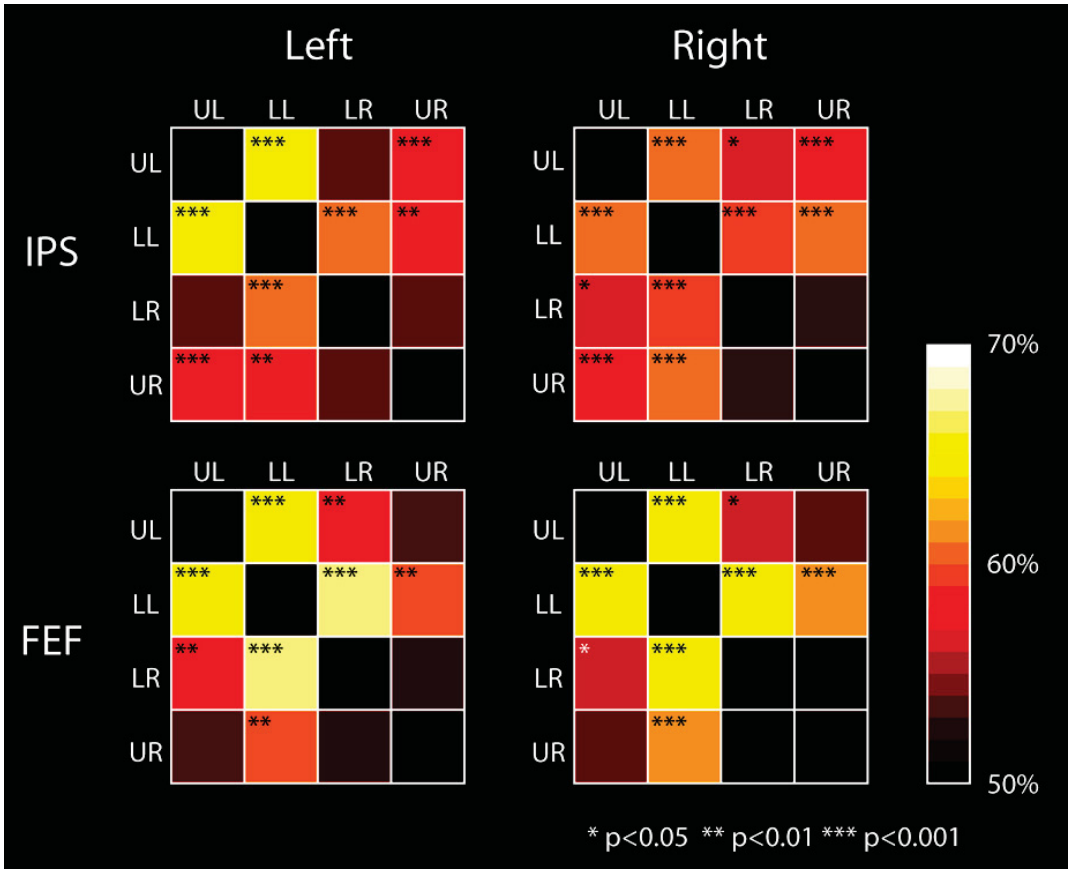

Figure S2. Results of All Six Possible Pairwise Searchlight Multivariate Pattern Classification Analyses for all Four Quadrants in Right and Left IPS and FEF

Chance level was 50% for these analyses. UL=upper left, LL=lower left, LR=lower right, UR=upper right quadrant (UL vs. LL, UL vs. LR, UL vs. UR, LL vs. LR, LL vs. UR and LR vs. UR). Regions in which 4-class classification was possible (see main text) were additionally used as ROIs and decoding accuracies for the various pair-wise searchlight analyses were extracted and averaged. One-class classification (the diagonal of the matrix) was not meaningful and therefore not performed but set to chance level. Note that the matrices are symmetric. Above chance decoding accuracy in the 4-class classification analysis could in principle be driven by only one well decodable quadrant, leaving open if full saliency information was encoded in these regions. However, using all pair-wise decoding analyses demonstrated information about at least three of the quadrants encoded in bilateral IPS and FEF. Only UR vs. LR could not be distinguished.

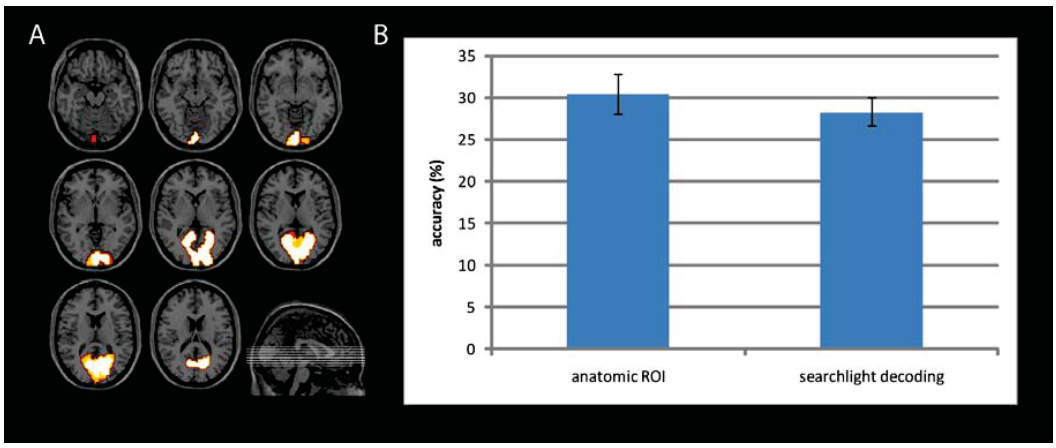

Figure S3.

(A) Region of interest (ROI) of the visual cortex along the calcarine sulcus. Data from this large ROI were used to perform the decoding analysis.

(B) Decoding performance for WTA thresholded saliency in a ROI defined from anatomy (calcarine sulcus) vs. averaged searchlight decoding performance in voxels that represented graded saliency in visual cortex ($t(20)=1.15$, $p=0.26$). Error bars represent standard error.

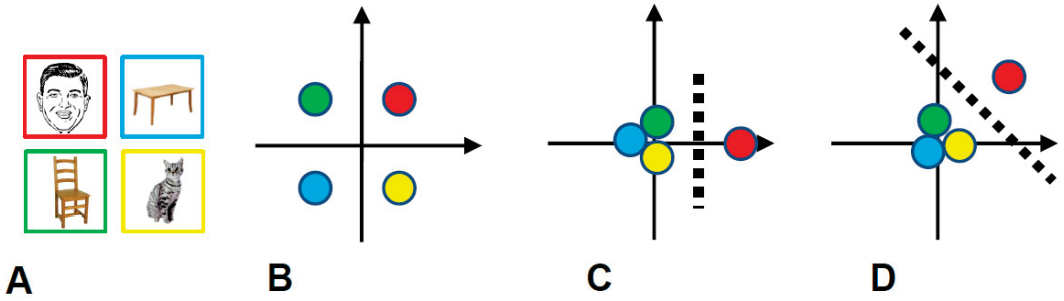

Figure S4.

The retinal image can be used to decode whether a face is present in a visual scene or not, which can be done using a face classification algorithm that could also be applied to bitmap images. However, the representation on the retina is usually not considered an explicit representation of a face, whereas a representation in the fusiform face area (FFA) is. Why? The reason is that the states of the retina do not make faces "special". They don't selectively differentiate faces from non-faces, they don't encode faces and objects as more different than any other two objects (say a table and a chair). In contrast, the FFA codes faces as one type of thing and all other objects as an undifferentiated ensemble of things. Thus, the representations selectively highlight the "faceness". The key question for an explicit representation is which aspect of the data is more emphasized by being "informationally separated" from others.

(A) Four examples of images that are to be represented, one of which is a face.

(B) A two-dimensional neural activation space where the activation vectors of each image are clearly separable and identifiable with a classifier. The face is no more different from the others than any other image.

(C) Here the face is clearly separable and the other images are not differentiated. This would count as an explicit coding of the "faceness" of the image. In this case the face can be identified by looking at one single dimension of neural activity (the x-axis).

(D) A similar case with an explicit face code, but this time the identification of the face is best when two dimensions are taken into consideration. Again, this is an explicit code, but the representation is *multivariate* instead of univariate. Please note that both of these representations, univariate and multivariate, can be easily read out from the next level of processing using a simple (neurophysiologically plausible) classifier. Thus, would be perfectly suitable for controlling eye movements.

Table S1. Regions where Graded Saliency in Any Quadrant Explained a Significant Proportion of Variance (Peak Positions Reported Only), Related to Figure 3

Anatomical Area	Visual Area	L/R	F-Value	X	Y	Z
Inferior occipital gyrus	V3v, V5, V4	L	23.47	-42	-81	-3
Middle occipital gyrus	V2, V1	L	10.47	-18	-96	12
	V5	R	19.13	45	-84	3
Occipital gyrus		R	14.28	33	-69	33
Fusiform gyrus		L	16.75	-36	-42	-24
		L	13.01	-45	-57	-21
		R	11.88	42	-45	-18
	V3v, V4	L	11.68	-21	-84	-9
Lingual gyrus	V2, V1	R	12.61	9	-84	-6
Superior occipital gyrus	V2, V1	L	10.95	-15	-102	12
Superior Parietel Lobule		L	11.32	-21	-60	51
Area 17		L	11.13	-36	-66	9

The coordinates are given according to Montreal Neurological Institute (MNI) space with their F-values (based on an ANOVA with one regressor per quadrant). L = left hemisphere, R = right hemisphere; all FWE-corrected ($p < .05$). The identification of visual areas is based on information from a standard cytoarchitectonic atlas [1].

Table S2: Regions where Saliency in One Quadrant Explained Significantly More Variance Than in Any Other (Peak Positions Reported Only), Related to Figure 4

Anatomical Area	Visual Area	L/R	F-Value	X	Y	Z
Fusiform gyrus	V3v, V4	L	9.95	-21	-84	-9
Lingual gyrus	V2, V1	R	11.70	9	-87	-6
Superior occipital gyrus	V2	L	10.11	-9	-99	21
Middle cingulate cortex		L	8.31	-9	-27	45
Caudate nucleus		L	9.56	-33	-6	24
Parietal operculum		L	8.69	-30	-27	24

The coordinates are given according to MNI space with their F-values. L = left hemisphere, R = right hemisphere; all $p < .0005$ uncorrected, voxel threshold = 10 voxels (based on an ANOVA with one regressor per quadrant). The identification of visual areas is based on information from a standard cytoarchitectonic atlas [1].

Table S3: Regions with Peak Information Regarding the Most Salient Quadrant (WTA), Related to Figure 3

Anatomical Area	L/R	T-Value	X	Y	Z
Frontal eye field	L	7.23	-27	-3	36
	R	10.57	27	-9	51
Anterior IPS	L	7.95	-24	-48	51
	R	8.02	36	-42	54
Superior occipital gyrus	R	7.50	24	-75	15
Middle temporal gyrus	R	7.01	63	-30	-6
Insula	L	7.63	-24	18	9

The coordinates are given according to MNI space with their T-values. L = left hemisphere, R = right hemisphere; all FWE-corrected ($p < .05$).

Supplemental Experimental Procedures

Subjects

Twelve male and ten female subjects (mean age 25.4, range 22-32 years) took part in the study and gave written informed consent to the test procedure. The experiment was approved by the local ethics committee and was conducted according to the Declaration of Helsinki. All subjects were right-handed and had normal or corrected to normal visual acuity. Behavioral data from 5 subjects were not correctly recorded due to a technical failure of the response recording device. Based on online monitoring of their performance during scanning, however, it could be concluded that they permanently attended to the fixation task. One different subject showed poor performance on the fixation task (47.6 % misses) and subsequently reported problems to concentrate; this subject's data was excluded from all analyses. The remaining 16 subjects performed well on the challenging fixation task with an average of 87.4 % (SD 8.45) correct responses.

Visual Stimuli

We used 100 gray-scale photographs of natural scenes [2]. The photographs were taken in several zoos in Southern Germany. The high resolution (4064 x 2704) photographs were cropped to parts of 1024 x 768 and centred at a random position in order to remove the centering bias of the original versions. For the experiment, stimuli were scaled to 800x600 pixels and presented via a projector (resolution 1024x768 pixel, 60Hz) that projected from the head-end of the scanner onto a screen. Subjects viewed the projection through a mirror fixed on the head coil. The visual angle of the full images was $\alpha = 20.9°x16.1°$. The photographs were presented for 1 s with a pseudo-randomized inter-stimulus-interval (ISI) of 1.4, 3.8 or 6.2 s. Each 1-s presentation consisted of a photograph being flashed ON–OFF–ON–OFF–ON, with ON-phases corresponding to image presentation (200 ms) and OFF-phases corresponding to the presentation of a gray background (200 ms). Each image was presented once per run, resulting in 100 trials. Additionally, in each run we presented 30 null-events during which only the gray background was shown. Combining a short jittered ISI with the presentation of null-events allowed for optimal estimation of the hemodynamic response function (HRF) [3, 4, 5].

Procedure

Subjects were required to direct their attention to the center of the screen where a small white outline square (visual angle $\alpha = 0.3°x0.3°$) was superimposed on the photographs or the gray background. Subjects had to solve a demanding visual fixation task. Every 1200 ms, the square's left or right bar was removed for another 600 ms. Subjects had to indicate by button press whether the square opened up to the left or the right hand side. Responses were given with the index and middle finger of the right hand, using a response box. To familiarize the subjects with the speed of the task, each run started with 12 s fixation task without further visual stimulation. Stimuli were presented and responses were recorded with MATLAB 7.0 (The MathWorks, Inc.) in combination with the Cogent toolbox (http://www.vislab.ucl.ac.uk/Cogent). During each scanning session, 5 runs of the experiment were conducted (each run lasted 10 min. and 38.4 s).

Outside the scanner a behavioural control experiment was conducted. 19 subjects performed exactly the same task as in the scanner. After about 1.5 minutes the presentation terminated with a last photograph followed by a noise mask to cancel any visual persistence. Immediately after the presentation of the mask subjects were asked to indicate the most salient, most interesting, or most distracting quadrant. Statistical analyses revealed that the

number of hits was not different from chance ($Chi^2(1, N=19) = 0.44$, $p=0.51$). In debriefing interviews, most subjects reported that their judgment was purely based on guessing. They also could not describe the last photograph and confirmed that the fixation task was very hard.

Quadrant-Specific Saliency Estimation

For each of the 100 photographs the distribution of saliency across the visual field was computed using the SaliencyToolbox for Matlab [6] that calculates the saliency for any given input image based on a standard computational model of visual saliency [7, 8]. Computational saliency map models from different authors [9, 10, 11] are comparable in their predictive power for saccades, thus the different saliency maps are highly correlated. Subsequently, the saliency maps were divided into four quadrants along the horizontal and vertical meridians. For each photograph presented, an average saliency value was calculated for each quadrant separately. This was done by averaging all finer-scaled local saliency values within each quadrant as given by the initial saliency map. The central regions of the images (<4.2°) where the fixation task was presented were excluded from this analyses. In a next step we identified the quadrant with the highest saliency for each photograph. According to the saliency model this quadrant would also be selected by a WTA mechanism. Point-biserial correlation between the graded saliency and the WTA thresholded saliency was r=0.68, equivalent to an explained variance of 0.46.

Functional Imaging

A Siemens TRIO 3T scanner with standard head coil was used to acquire gradient-echo EPI functional MRI volumes covering the occipital and parietal lobe (36 axial slices, TR = 2400 ms, echo time TE = 30 ms, resolution 3x3x2 mm^3 with 1 mm gap). In each run, 266 images were acquired for each subject. The first five images were discarded to allow for magnetic saturation effects. Two different models were estimated for each subject in order to identify the neural substrate of the *graded saliency* and the output of a *winner-take-all threshold*.

Encoding of Graded Saliency

First, a general linear model (GLM) with 1 event-based and HRF-convolved regressor was estimated, separately for each voxel. For this the fMRI data were first motion corrected and then spatially smoothed with a Gaussian kernel of 6 mm FWHM using SPM2 (http://www.fil.ion.ucl.ac.uk/spm/). Data were highpass filtered with a cutoff period of 128 s. The first regressor of the GLM estimated the response to the onset of the stimuli. Additional four regressors were used to estimate the parametric modulation of this onset response by the saliency in each of the four quadrants. The resultant contrast maps were normalized to a standard stereotaxic space (Montreal Neurological Institute EPI template) and re-sampled to an isotropic spatial resolution of 3 x 3 x 3 mm^3. Finally, random effects general linear models were estimated across subjects. Regions encoding graded saliency were identified using an F-test based on a second-level ANOVA including the four parametric regressors.

In order to maximize the similarity between methodological approaches for the investigation of graded saliency and WTA-thresholded saliency (see below) additional multivariate pattern analysis was performed to search for regions where **distributed** local voxel ensembles encoded graded saliency. For this the averaged graded saliency values of all 100 stimuli were split into quartiles for each quadrant separately. A GLM was estimated for each of the quadrants and each of the 4 graded levels of saliency, resulting in 16 conditions in total. The GLM was based on motion corrected, non-normalized and unsmoothed data to

maximize the sensitivity for information encoded in fine-grained spatial voxel patterns [12, 13, 14], for a discussion see [15, 16, 17]. In order to estimate the information encoded in spatially distributed response patterns at each brain location, we employed a "searchlight" approach [18, 19, 20, 21] that allowed the unbiased search for informative voxels across the whole brain. A spherical cluster of N surrounding voxels ($c_{1...N}$) within a radius of six voxels was created around a voxel v_i. The GLM-parameter estimates for these voxels were extracted and transformed into vectors for each condition for each run of each subject. These vectors represented the average spatial response patterns [22] to the given condition from the chosen cluster of voxels. In the next step, multivariate pattern regression was used to assess whether information about the 4 levels of saliency was encoded in the spatial response patterns. For this purpose, the pattern vectors from four of the five runs were assigned to a "training data set" that was used by a radial basis function (RBF) support vector pattern regression [23] with a fixed regularisation parameter C = 1 and γ = 1/number of features for the RBF. First, the support vector regression was trained on this data to identify patterns corresponding to each of the four levels of saliency (LIBSVM implementation, http://www.csie.ntu.edu.tw/~cjlin/libsvm). Then it predicted independent data from the last run ("test data set"). Cross-validation (5-fold) was achieved by repeating this procedure independently, with each run acting as the test data set once, while the other runs were used as training data sets. This procedure prevented overfitting and "double dipping" [24]. The correlation between the predicted and real saliency levels was Fisher-z normalized, averaged across all five iterations and assigned to the central voxel v_i of the cluster. It therefore reflected the fit of the regression based on the given spatial activation patterns of this local cluster. A correlation significantly above zero implied that the local cluster of voxels spatially encoded information about the saliency level of one quadrant, whereas a correlation of zero implied no information. The same analysis was then repeated with the next spherical cluster, created around the next spatial position at voxel v_j. Again, an average correlation for this cluster was extracted and assigned to the central voxel v_j. By repeating this procedure for every voxel in the brain, a 3-dimensional map of correlation coefficients for each position could be created. The analysis was performed for each quadrant separately. The resultant correlation maps were normalized to a standard stereotaxic space (Montreal Neurological Institute EPI template), re-sampled to an isotropic spatial resolution of 3 x 3 x 3 mm^3 and smoothed with a Gaussian kernel of 6 mm FWHM using SPM2. Finally, a random effects analysis was conducted, computed on a voxel-by-voxel basis, to statistically test the correlation for each position in the brain across all subjects [19].

WTA-Thresholded Saliency

For the analysis of WTA-thresholded saliency we estimated a GLM with 4 event-based and HRF-convolved regressors, each representing time points where one quadrant was maximally salient. These will be referred to as the four conditions of the experiment. The GLM for the WTA-outcome was based on motion corrected, non-normalized and unsmoothed data to maximize the sensitivity for information encoded in fine-grained spatial voxel patterns [12, 13, 14], for a discussion see [15, 16, 17]. In order to estimate the information encoded in spatially distributed response patterns at each brain location, we employed a "searchlight" approach [18, 19, 20, 21] that allowed the unbiased search for informative voxels across the whole brain. A spherical cluster of N surrounding voxels ($c_{1...N}$) within a radius of six voxels was created around a voxel v_i. The GLM-parameter estimates for these voxels were extracted and transformed into vectors for each condition for each run of each subject. These vectors represented the average spatial response patterns [22] to the given condition from the chosen cluster of voxels. In the next step, multivariate pattern classification was used to assess

whether information about the experimental condition was encoded in the spatial response patterns. For this purpose, the pattern vectors of four of the five runs were assigned to a "training data set" that was used by a radial basis function (RBF) support vector pattern classifier [23] with a fixed regularisation parameter C = 1 and γ = 1/number of features for the RBF. First, the classifier was trained on this data to identify patterns corresponding to each of the four conditions (LIBSVM implementation, http://www.csie.ntu.edu.tw/~cjlin/libsvm). Then it classified independent data from the last run ("test data set"). The multiclass classification was achieved by combining several pair-wise SVMs. Cross-validation (5-fold) was done by repeating this procedure independently, with each run acting as the test data set once, while the other runs were used as training data sets. This procedure prevented overfitting and "double dipping" [24]. The decoding accuracy was assessed by averaging the results of all five classification iterations and was assigned to the central voxel v_i of the cluster. It therefore reflected the accuracy of classification based on the given spatial activation patterns of this local cluster. Classification accuracy significantly above chance (25% for four quadrants) implied that the local cluster of voxels spatially encoded information about the quadrant, whereas chance level performance implied no information. The same analysis was then repeated with the next spherical cluster, created around the next spatial position at voxel v_j. Again, an average decoding accuracy for this cluster was extracted and assigned to the central voxel v_j. By repeating this procedure for every voxel in the brain, a 3-dimensional map of decoding accuracies for each position could be created. The resultant accuracy maps were normalized to a standard stereotaxic space (Montreal Neurological Institute EPI template), re-sampled to an isotropic spatial resolution of 3 x 3 x 3 mm^3 and smoothed with a Gaussian kernel of 6 mm FWHM using SPM2. Finally, a random effect analysis was conducted, computed on a voxel-by-voxel basis, to test the decoding accuracy statistically for each position in the brain across all subjects [19].

WTA-Thresholded Saliency in Primary Visual Cortex

We constructed a region of interest (ROI) of early visual cortex around the calcarine sulcus, covering V1, based on the Anatomical Automatic Labeling toolbox for SPM [25] (see Figure S3). This large ROI covers the representation of the whole stimulated visual field in V1. Multivariate pattern classification for the most salient quadrant was then calculated with identical parameters as in the searchlight decoding analysis (see above). We then compared the decoding performance for the large early visual cortex ROI with decoding performance of the averaged searchlight decoding in regions in which graded saliency was represented. If receptive field size or classifier size restricted our findings, this ROI decoding analysis should clearly outperform the original analysis. The comparison revealed, however, that there was no difference ($t(20)=1.15$, $p=0.26$; see Figure S3) in decoding accuracy between both approaches and thus our results are not due to the restricted size of our searchlights.

Supplemental References

1. Amunts, K., Schleicher, A., and Zilles, K. (2007). Cytoarchitecture of the cerebral cortex–more than localization. Neuroimage, *37*, 1061–5.
2. Wichmann, F., Kienzle, W., Schölkopf, B., and Franz, M. (2008). Visual saliency re-visited: Center-surround patterns emerge as optimal predictors for human fixation targets. J. Vis., *8*, 635–635.
3. Burock, M.A., Buckner, R.L., Woldorff, M.G., Rosen, B.R., and Dale, A.M. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional mri. Neuroreport, *9*, 3735–3739.

4. Dale, A.M. (1999). Optimal experimental design for event-related fmri. Hum. Brain Mapp., *8*, 109–114.
5. Friston, K.J., Zarahn, E., Josephs, O., Henson, R.N., and Dale, A.M. (1999). Stochastic designs in event-related fmri. Neuroimage, *10*, 607–619.
6. Walther, D. and Koch, C. (2006). Modeling attention to salient proto-objects. Neural Netw., *19*, 1395–1407.
7. Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol., *4*, 219–227.
8. Itti, L. and Koch, C. (2001). Computational modelling of visual attention. Nat. Rev. Neurosci., *2*, 194–203.
9. Gao, D., Mahadevan, V., and Vasconcelos, N. (2008). On the plausibility of the discriminant center-surround hypothesis for visual saliency. J. Vis., *8*, 13.1–1318.
10. Bruce, N.D.B. and Tsotsos, J.K. (2009). Saliency, attention, and visual search: An information theoretic approach. J. Vis., *9*, 1–24.
11. Kienzle, W., Franz, M.O., Schölkopf, B., and Wichmann, F.A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. J. Vis., *9*, 1–15.
12. Kamitani, Y. and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci., *8*, 679–685.
13. Haynes, J.D. and Rees, G. (2006). Decoding mental states from brain activity in humans. Nat. Rev. Neurosci., *7*, 523–534.
14. Haynes, J.D. and Rees, G. (2005). Predicting the stream of consciousness from activity in human visual cortex. Curr. Biol., *15*, 1301–1307.
15. de Beeck, H.P.O. (2010). Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fmri analyses? Neuroimage, *49*, 1943–1948.
16. Kamitani, Y. and Sawahata, Y. (2010). Spatial smoothing hurts localization but not information: pitfalls for brain mappers. Neuroimage, *49*, 1949–1952.
17. Swisher, J.D., Gatenby, J.C., Gore, J.C., Wolfe, B.A., Moon, C.H., Kim, S.G., and Tong, F. (2010). Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. J. Neurosci., *30*, 325–330.
18. Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. Proc. Natl. Acad. Sci. U S A, *103*, 3863–3868.
19. Haynes, J.D., Sakai, K., Rees, G., Gilbert, S., Frith, C., and Passingham, R.E. (2007). Reading hidden intentions in the human brain. Curr. Biol., *17*, 323–328.
20. Soon, C.S., Brass, M., Heinze, H.J., and Haynes, J.D. (2008). Unconscious determinants of free decisions in the human brain. Nat. Neurosci., *11*, 543–545.
21. Bode, S. and Haynes, J.D. (2009). Decoding sequential stages of task preparation in the human brain. Neuroimage, *45*, 606–613.
22. Mourão-Miranda, J., Reynaud, E., McGlone, F., Calvert, G., and Brammer, M. (2006). The impact of temporal compression and space selection on svm analysis of single-subject and multi-subject fmri data. Neuroimage, *33*, 1055–1065.
23. Müller, K.R., Mika, S., Ratsch, G., Tsuda, K., and Scholkopf, B. (2001). An introduction to kernel-based learning algorithms. IEEE Trans. Neural Netw., *12*, 181–201.
24. Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., and Baker, C.I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci., *12*, 535–540.
25. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. Neuroimage, *15*, 273–289.

APPENDIX

A Publikationen

Forschungsartikel

Betz T., Wilming N., **Bogler C.**, Haynes J.D., König P. (submitted to *Journal of Vision*). Dissociation between saliency signals and activity in early visual cortex.

Bogler C., Bode S., Haynes J.D. (2013). Orientation pop-out processing in human visual cortex. *Neuroimage,* 81: 73-80.

Bode, S., **Bogler, C.**, Haynes, J.D. (2013). Similar neural mechanisms for perceptual guesses and free decisions. *Neuroimage*, 65, 456-65

Bode, S., **Bogler, C.**, Soon, C.S., Haynes, J.D. (2012). The neural encoding of guesses in the human brain. *Neuroimage*, 59(2), 1924-31

Bogler, C., Bode, S., Haynes, J.D. (2011). Decoding successive computational stages of saliency processing. *Current Biology*, 21(19), 1667-71.

Ferstl, E.C., Neumann, J., **Bogler, C.**, von Cramon, D.Y. (2008). The extended language network: a meta analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, 29(5), 581-93

Konferenzbeiträge (Auswahl)

Bogler, C., Kuhlen, K.K., Swerts, M.G.,, S. Haynes, J.D. Neuronal coding of assessing another person's knowledge based on nonverbal cues. Vortrag bei der Organization for Human Brain Mapping Annual Meeting (*OHBM*) 2013

Bogler, C., Mehnert, J., Steinbrink, J., Haynes, J.D. Decoding vigilance from NIRS-data: NIRS for BCI applications? Poster bei der Organization for Human Brain Mapping Annual Meeting (*OHBM*) 2011

Bogler, C., Bode, S. Haynes, J.D. Multivariate decoding reveals successive computational stages of saliency processing. Poster bei der Organization for Human Brain Mapping Annual Meeting (*OHBM*) 2009

Bogler, C. & Haynes, J.D. Retinotopically independent processing of saliency signals in the near-absense of attention. Vortrag beim International Congress of Psychology (*ICP*) 2008

Bogler, C. & Haynes, J.D. Retinotopically independent processing of saliency signals in the near-absence of attention. Poster bei der Konferenz der Vision Sciences Society (*VSS*) 2008

Bogler, C. & Haynes, J.D. Retinotopically independent processing of saliency signals in the near-absence of attention. Poster beim Bernstein Symposium 2008

Bogler, C. & Haynes, J.D. Retinotopically independent processing of saliency signals in the near-absence of attention. Poster bei der Organization for Human Brain Mapping Annual Meeting (*OHBM*) 2007

B Selbständigkeitserklärung

Hiermit versichere ich, dass ich die Dissertation selbständig und ohne unerlaubte Hilfe angefertigt habe. Ich habe die Dissertation an keiner anderen Universität eingereicht und besitze keinen Doktorgrad im Fach Psychologie. Die Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät II vom 17.01.2005, zuletzt gändert am 13.02.2006, veröffentlicht im Amtlichen Mitteilungsblatt Nr. 34/2006 ist mir bekannt.

_____

Ort, Datum Unterschrift